FEB 28 1961

March, 1961
Vol. 32, No. 1

Child Development

CONTENTS

Developmental Changes in the Effects of Reinforcement and Nonreinforcement of a Single Response
Harold W. Stevenson and Morton W. Weir 1

Selection and Test of Response Measures in the Study of the Human Newborn
William Kessen, E. Jane Williams, and Joanna P. Williams 7

Form, Color, and Size in Children's Conceptual Behavior
Jerome Kagan and Judith Lemkin 25

Developmental Changes in Preference for Goals Difficult to Attain
Donald C. Smith and Lucy Wing 29

Effect of Exposure to Symbolic Aggression on Aggressive Behavior
O. Ivar Lövaas 37

Sex Differences in Aggression and Its Correlates in Middle-Class Adolescents
Leonard M. Lansky, Vaughn J. Crandall, Jerome Kagan, and Charles T. Baker 45

Experimental Development of "Like" and "Dislike" of Others among Adolescent Girls
Evan R. Keislar 59

Effect of Withdrawal of Positive Reinforcement on an Extinguishing Response in Young Children
Donald M. Baer 67

An Exploration of Patterns of Impulse Control in Early Adolescence
Norman Livson and Wanda C. Bronson 75

Level of Stimulus Pretraining and Paired-Associate Learning
Hayne W. Reese 89

Measurement of Movement in the Human Newborn: a New Technique
William Kessen, Louise S. Hendry, and Anne-Marie Leutzendorff 95

Amplitude of an Instrumental Response Following the Cessation of Reward
Ruth B. Holton 107

Relative Effectiveness of Different Reinforcement Combinations on Concept Learning of Children at Two Developmental Levels
William J. Meyer and Stanley B. Seidman 117

Normative Data on the Children's Manifest Anxiety Scale at the Rural Third Grade Level
Harold D. Holloway 129

Social Histories of Successful and Unsuccessful Children
John W. Baker and Annette Holzworth 135

(continued on back cover)

Published Quarterly by Child Development Publications of the

SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, INC.

Editor

WILLIAM E. MARTIN
Purdue University

Consulting Editors

DONALD M. BAER
University of Washington

EMANUEL K. BELLER
Child Development Center

ALFRED CASTANEDA
University of Texas

WALTER EMMERICH
University of Colorado

JEROME KAGAN
Fels Research Institute

IRVIN A. KRAFT
Houston State Psychiatric Institute

LEWIS P. LIPSITT
Brown University

NORMAN LIVSON
University of California

JAMES M. TANNER
University of London

RICHARD H. WALTERS
University of Toronto

Managing Editor

KATE HOFFMAN

Child Development, one of three publications of the Society for Research in Child Development, Inc., is issued quarterly in March, June, September, and December. The subscription price per year is $12.00 domestic, $12.50 foreign. Single issues are $3.50.

Child Development Abstracts and Bibliography is issued three times a year, two numbers in each issue. The subscription price per year is $6.00 domestic, $6.50 foreign. Single issues are $2.50.

Monographs of the Society for Research in Child Development is issued at irregular intervals during the year. The subscription price per year is $6.00 domestic, $6.50 foreign. The price of individual monographs is variable.

The three publications of the Society are available at the special price of $20.00 per year, domestic; $21.50 per year, foreign.

A limited number of back issues of all publications is available. A mimeographed list may be had upon request.

Subscriptions, address changes, and business communications should be sent to CHILD DEVELOPMENT PUBLICATIONS, Purdue University, Lafayette, Indiana.

Please give four weeks' notice when changing your address, giving old and new addresses. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by
THE ANTIOCH PRESS
Yellow Springs, Ohio

Child Development

VOLUME 32

MARCH, 1961 *to* DECEMBER, 1961

Editor

WILLIAM E. MARTIN
Purdue University

Consulting Editors

DONALD M. BAER
University of Washington

EMANUEL K. BELLER
Child Development Center

ALFRED CASTANEDA
University of Texas

WALTER EMMERICH
University of Colorado

JEROME KAGAN
Fels Research Institute

IRVIN A. KRAFT
Houston State Psychiatric Institute

LEWIS P. LIPSITT
Brown University

NORMAN LIVSON
University of California

JAMES M. TANNER
University of London

RICHARD H. WALTERS
University of Toronto

Managing Editor

KATE HOFFMAN

CHILD DEVELOPMENT PUBLICATIONS

Society for Research in Child Development, Inc.
Purdue University, Lafayette, Indiana

1961

DEVELOPMENTAL CHANGES IN THE EFFECTS OF REINFORCEMENT AND NONREINFORCEMENT OF A SINGLE RESPONSE [1]

HAROLD W. STEVENSON* *and* MORTON W. WEIR
University of Minnesota

This study investigates developmental changes in the effects of reinforcement and nonreinforcement of a response in a three-choice situation. The responses of preverbal children in learning situations (e.g., 3, 4) appear to be in line with predictions from reinforcement theories of behavior. Reinforcement of a response appears to increase the frequency with which the response is repeated and nonreinforcement to decrease this frequency. Older children and adults, however, have been found to behave in a different fashion. For example, Jarvik (2) and Gratch (1) have found evidence for a "negative recency effect" in these *Ss*' responses; that is, these *Ss* tend to predict that the previously occurring or reinforced event will not recur immediately and select an alternative event on a subsequent trial. There appear to be developmental changes in the effects of reinforcement and nonreinforcement on behavior.

The present investigation utilizes data collected in previously published studies (5, 6). A three choice-situation was employed in which one of the choices was reinforced and the other two choices were not. The analysis of behavior during the first few trials in this task makes it possible to determine the effects of initial reinforcement or nonreinforcement of a response on the tendency of *Ss* at different development levels to repeat this response on the subsequent trial.

METHOD

Subjects

The records of 380 children were reviewed to obtain (a) 30 *Ss* at each age level who were not reinforced on their initial trial and (b) 30 *Ss* at each age level who received reinforcement at the latest on their fourth choice

* Institute of Child Development and Welfare, University of Minnesota, Minneapolis 14.

[1] The study was supported by a grant from the National Science Foundation (Grant 3280) to the senior author.

of the reinforcing stimulus. Children at four age levels were selected: 3-0 to 3-11, 5-0 to 5-11, 7-0 to 7-11, and 9-0 to 9-11 years.

Apparatus

The apparatus has been described previously (6). It consisted of a yellow panel containing a horizontal row of three knobs. Above the knobs was a signal light and below the knobs was a delivery hole for marbles. The marbles fell into an enclosed plastic container.

Procedure

An identical procedure was followed in the studies from which the data for this investigation were collected. Essentially, it consisted of *S*'s being instructed to push one of the three knobs when the signal light came on and being told that if he pushed the correct knob a marble would fall into the plastic container. One of the three knobs was designated the correct knob for a particular *S* and, depending upon the condition under which *S* was tested, either 33, 66, or 100 per cent of the choices of this knob were reinforced. The other two knobs never yielded reinforcement.

Results

The record of each *S* was examined to determine the response of *S* on the subsequent trial when (a) *S*'s initial response in the task was not reinforced, (b) *S*'s initial response to the reinforcing stimulus was reinforced, and (c) *S*'s response to the reinforcing stimulus was reinforced for the first time. The latter two conditions differ in that in (b) *S*'s first response to the reinforcing stimulus was reinforced, while in (c) *S*'s initial response to the reinforcing stimulus was not reinforced but a later response was reinforced. No more than three nonreinforced responses to the reinforcing stimulus occurred in the latter case prior to the first reinforced trial. The subsequent response was classified according to whether or not the prior response was repeated by *S*.

Effects of Nonreinforcement

Performance on trial 2 was examined for *S*s who were not reinforced for their response on trial 1. The proportion of *S*s at each CA level who changed response on trial 2 is presented in Figure 1. The *S*s at CA 3 show some tendency to repeat the nonreinforced response, while very few *S*s at the older age levels show this tendency. The difference between the performance of *S*s at CA 3 and the other groups combined is highly significant ($\chi^2 = 24.87$, $df = 1$, $p < .001$).

Effects of Reinforcement

It is possible that performance on the subsequent trial might differ for *S*s who were reinforced on their initial choice of the reinforcing stimulus and for *S*s who were not reinforced for their choices of the reinforcing

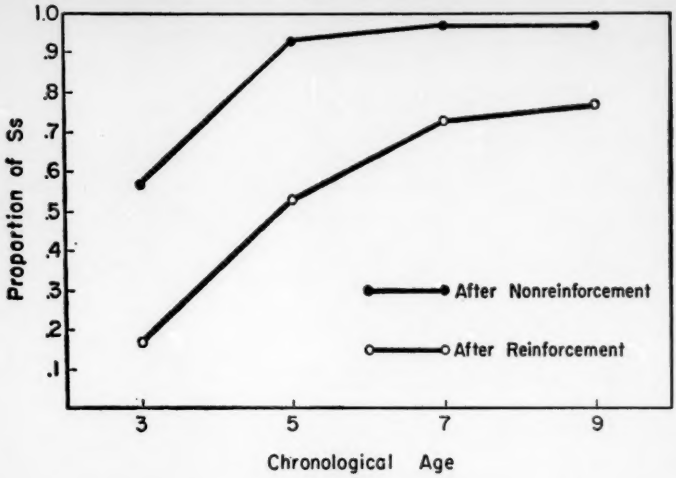

FIGURE 1—Proportion of *Ss* at each CA level changing response following reinforcement and nonreinforcement of the prior response.

stimulus on several trials prior to receiving their first reinforcement. In order to determine whether these two conditions differ, a comparison was made of the *Ss* in situations (b) and (c) described above. Of 40 *Ss* (10 at each CA level) who were reinforced on their first choice of the reinforcing knob, 22 repeated the response on the next trial, while of 80 *Ss* (20 at each CA level) who received reinforcement only after choosing the reinforcing stimulus up to three times with no reinforcement, 32 repeated the response on the next trial. The difference between these two conditions is not significant ($\chi^2 = 1.86$, $df = 1$, $p > .10$), and for further analysis the *Ss* in the two situations were combined.

The proportion of *Ss* at each CA level who changed response following reinforcement is presented in Figure 1. A consistent increase through the four age levels is seen in the number of *Ss* changing response. The increase with age is more gradual than that found in the analysis of the effects of nonreinforcement. The results for the older *Ss* provide clear evidence for the operation of a negative recency effect. The differences among the four age groups are highly reliable ($\chi^2 = 27.61$, $df = 3$, $p < .001$), indicating significant changes in the effects of reinforcement as a function of CA.

Data for Adult Subjects

It is of interest to compare the performance of the older children with that of adults (college students), for whom data were available in the same

situation under the same experimental conditions. The number of *Ss* changing response following their first reinforced trial continued to increase from the 9-year to the adult level. Of the 30 adult *Ss*, 90 per cent changed response on the subsequent trial. Comparing these results with those in Figure 1, it can be seen that after the age of 7 there is relatively little change in the tendency of individuals to change response after their initial reinforced trial and that the negative recency effect is manifest in the performance of most *Ss*. The comparison of the child and adult data for trial 2 following nonreinforcement on trial 1 does not show a consistent developmental change. Ten of the 30 adult *Ss* and only one of the 30 9-year-old children repeated the nonreinforced response on trial 2 ($\chi^2 = 7.13$, $df = 1$, $p < .01$).

Discussion

The results of these analyses provide clear evidence that the effect of reinforcement and nonreinforcement of a single response differs significantly as a function of chronological age. The behavior of the 3-year-olds is readily accounted for by traditional S-R concepts. Reinforcement increased the tendency to repeat the reinforced response on the subsequent trial and nonreinforcement reduced this tendency. By CA 7, reinforcement of a response did not increase the tendency to repeat the response; in fact, the effect appeared to be just the opposite, for *Ss* tended to change their response following reinforcement to nearly as great a degree as they did following nonreinforcement.

These results do not, of course, indicate that an S-R analysis of the behavior of older children is impossible. They do indicate that some other unit of response must be utilized in discussing older *Ss*' behavior. It may be assumed that the behavior of older children is guided by verbal hypotheses and that it is in the confirmation of these hyopotheses that reinforcement has its greatest effect. Children's verbalizations of their basis of response seem to indicate that one of the dominant hypotheses in a multiple-response situation is that there is a pattern of responses which leads to reinforcement. An analysis of behavior in terms of patterns or sequences of responses may indicate that nonreinforcement and reinforcement have the same effect on these units of response for older *Ss* as they do on the single responses of younger *Ss*.

Summary

Data for 30 *Ss* at each of four age levels from 3 through 9 years were selected to determine the effects of initial nonreinforcement and reinforcement of a single response on behavior during the subsequent trial. Significant differences were found in the effects of both reinforcement and nonreinforcement as a function of CA. A strong tendency was found for the negative recency effect to increase with increasing CA. The results were

interpreted as indicating that utilization of a single S-R unit is inappropriate in the analysis of the behavior of older children and adults.

References

1. Gratch, G. The development of the expectation of the nonindependence of random events in children. *Child Develpm.*, 1959, 30, 217-228.
2. Jarvik, M. H. Probability learning and a negative recency effect in the serial anticipation of alternating symbols. *J. exp. Psychol.*, 1951, 41, 291-297.
3. Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, 36, 471-490.
4. Siegel, P. S., & Foshee, J. G. The law of primary reinforcement in children. *J. exp. Psychol.*, 1953, 45, 12-15.
5. Stevenson, H. W., & Weir, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, 57, 403-412.
6. Stevenson, H. W., & Zigler, E. F. Probability learning in children. *J. exp. Psychol.*, 1958, 56, 185-192.

SELECTION AND TEST OF RESPONSE MEASURES IN THE STUDY OF THE HUMAN NEWBORN[1,2]

WILLIAM KESSEN,* E. JANE WILLIAMS, *and* JOANNA P. WILLIAMS
Yale University

Thirty years ago, Weiss stated the problems of response selection in studies of the human newborn:

> . . . There is very little coordination of the different movements. . . . The assumption that no movement is ever *repeated exactly* becomes a very obvious fact. It seems at first futile to speak of classifying the behavior. While this is true theoretically, as a matter of fact we do group together certain types of movements according to their resemblances (29, p. 454).

Unfortunately, Weiss's criterion of classification—the occurrence of "resemblances"—does not provide an unambiguous rule for ordering the behavior of newborns. Should we attend to resemblances in anatomical topography, or speed, or temporal overlap, or effect on the environment, or relation to antecedent stimulation? In fact, all of these criteria for classification have been used by students of neonatal behavior, with little evidence that a simple general rule for response definition is emerging. The problem of response definition is by no means confined to research with infants—it has long been an issue in general psychology (19, 22)—but the newborn presents special difficulties in answering this most basic of research questions: What are the effective response-classes?

In the observation of animals and older human beings, psychologists of all theoretical commitments depend on *movement in relation to the environment* and on *speech* to provide the most elemental definition of response categories. Whether in the study of maze-learning or of psychotherapy, the

* Department of Psychology and Child Study Center, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported here was supported in part by a grant from Fluid Research Fund, Yale University School of Medicine, and in part by Research Grant M-1787, United States Public Health Service.

[2] The writers are deeply indebted to Dr. Milton J. E. Senn and to the staff of Grace-New Haven Community Hospital for their generous cooperation. S. J. Onesti, J. Doris, J. C. Beck, A. L. S. Hendry, and A. M. Leutzendorff assisted in the collection and analysis of data during the research and in observations preliminary to it.

psychologist of mature behavior can make use of the fact that his subjects move through space or speak; the student of newborns can claim no such advantage. This difference turns out to be quite important in the first stages of the search for meaningful response categories. Although general definitions of response-classes have been developed (e.g., those responses which increase in probability of occurrence with the occurrence of reinforcement), a great deal of psychological observation still depends on the use of categories common to "natural" everyday usage. Thus, the behavior of picking up a cigaret and the behavior of picking up a rattle can "naturally" be put in the response-class Picking Up, even though the responses may differ along many dimensions. Similarly, the expression "I hate dishwashing" and the expression "Golf is repulsive to me" will naturally indicate the same response-class of Disliked Activities even though they are phonetically quite different and may differ on other grounds as well. People organized the activities of other people into response-clusters long before there were professional psychologists, and these vernacular rules for the classification of behavior seem to depend in large measure on characteristics of language and on regularities in response to objects (including other organisms) in the environment. When the psychologist sets out on his systematic study, he almost invariably adapts his observation to these everyday classificatory rules. Such a strategy is reasonably successful when used with mature human beings and even with animals, but in most cases of newborn behavior the "natural" rules are inapplicable. It is largely for this reason that the behavior of newborns is seen as meaningless; we are unable to organize the neonate's behavior with the rules of classification natural for older organisms.[3]

Curiously enough, the lack of obvious criteria of meaningfulness in newborn behavior also precludes exclusion of response-classes; that is, any behavior might be productively studied. This ambiguity, which is a problem for all psychologists but of special importance for those of us interested in early behavior, forces attention to the more general question: What do you want the response-categories for? There are several ways in which this question has been answered in the past, and a brief review of some of the positions taken will put the issue of response selection in a sharper light.

In the tradition of comparative psychology, the human newborn can be seen as a separate species, or at least a separate breed, of interest to the psychologist because of his variation from the behavior of other animals. Much of the classical research on neonatal behavior fits with this view of the infant. Can he discriminate color (27)? What are his taste sensitivities (17)? Can he learn (24)? Is he responsive to privation (16)? What reflexive behavior does he show (2)? Questions of this order, which led to the accumulation of important observations of infant behavior, were almost in-

[3] Wolf (30) has suggested that the basic natural rule for organizing the behavior of human beings depends on the observation of "goal" or "endpoint" behavior. If this is the case, then the apparent meaninglessness of the newborn's activities may represent the inability of adult observers to recognize the newborn's "goals."

variably derived from general psychology; those independent variables were manipulated which had been important in the study of other animals, and response definitions were developed which showed more-or-less reliable relationships with external stimulation. Dennis's (2) survey of the response repertory of the human infant lists the response-classes which resulted from this pioneer work on neonatal reflexology, and his summary is convincing demonstration that a relatively large number of response categories for neonatal behavior can be described under the criterion of close temporal relation to external stimulation.

In the course of early research on newborn behavior, there was developed a response-class very different from the simple "reflexive" activities which dominate Dennis's list. Called variously "mass activity" or "general activity," this response category was the antithesis of the stimulus-response descriptions and had as its criterion of inclusion *any* detectable movement of the newborn. Several techniques were used to determine measures for this response-class, but in general they aimed at representing all movements of the infant, whether these movements, collected in smaller subclasses, could be reliably related to external stimulation or not. For example, a response class defined in this way was the chief tool in Irwin's fundamental work describing the effect of several antecedent manipulations on neonatal activity (e.g., 15). There was promise, during the thirties, that "general activity" would serve for the study of newborn behavior the same broad function that "pressing a bar" or "running a maze" had served for other psychological problems. Perhaps because the specific measures devised for neonatal activity were not sensitive enough to detect response changes of brief duration, this promise was not fulfilled.

The study of the newborn as species, directed more by interest in reliable description than by commitment to elaborated theoretical issues, produced a natural history of neonatal behavior which was fundamental to later study and developed a response class—general activity—which was shown to be responsive to manipulation.

Running alongside research interest in newborns as behavioral variants, and never independent of it, has been the study of neonatal behavior as the precursor of later behavior. In this view, the newborn is seen as part of a long development toward adult human functioning, and response-classes have been sought which will show continuity of this development over time. However, as one might expect from our earlier discussion, there are no obvious and natural continuities between the behavior of newborns and the behavior of older human beings. As a result, the search for genotypic continuity in the face of phenotypic diversity (30) has led to a stronger emphasis on theory—that is, on postulated continuity—than was the case for the comparative psychologists of infancy. From this search for developmental laws have come the three most important contributions to a general theory of newborn behavior—Freud's theory of personality development (e.g., 3), Gesell's theory of maturation (e.g., 9), and Piaget's theory of the growth

of intelligence (e.g., 26). Each of these systems, devised to deal with the entire range of human development, makes statements about the important response-classes to be observed in the study of newborns.

Gesell's work has had by far the widest research impact, leading to a line of studies on the normative development of the human being, particularly with regard to relatively simple motor behavior (e.g., 10, 13). Piaget's interest has been in aspects of neonatal behavior which show the operation of what he considers general intellective functions, and his observations (26) have as a theme the infant's development of increasingly effective ways of dealing with the environment. To date, no systematic study has been made of newborn behavior with Piaget's postulations as guide. The developmental theory outlined by Freud, which emphasized the early expression and control of impulse, has produced a great deal of commentary (4, 12, 14) and some provocative research (e.g., 28). Its implications for the study of neonatal behavior have been treated in detail by Lustman (23), who also discusses the data on newborn behavior relevant to psychoanalytic theory.

Whatever theoretical position is taken, the study of the immature organism as precursor of the mature organism involves two interrelated problems: the establishment of developmental sequences common to all human beings (e.g., changes in prone progression, psychosexual development) and the determination of parameters which permit the description of individual differences (e.g., accelerated development, regression rate). In the specific instance of the newborn, one may ask: What behavior or change in behavior is common to all infants (e.g., occurrence of sucking movements when orally stimulated). Or one may ask: What behavior or change in behavior is characteristic of one child or of a limited group of children (e.g., variations in intensity of sucking movements). Although the first kind of normative question receives continuing examination, research interest in the study of newborns has shifted in recent years to the second. The investigations of Fries (5, 6) on activity, of Balint (1) on sucking style, of Lustman (23) on response to tactile stimulation, and of Graham *et al.* (11) on sensitivity, among others, reflect an increasing interest in the dimensions of variations among human newborns.

In short, the consideration of the newborn as a precursor brings to the problem of response selection two points of emphasis: the relevance of neonatal behavior to later behavior and the detection of individual differences. Attention to these aspects of the problem also makes clear that the lack of obvious and common-sense continuities between neonatal and later behavior requires that the selection of response-classes be directed in the first instance by hunch or preconception or theory.

Once a response-class has been selected for study, however, whether on the basis of a specific developmental theory or on the basis of a less well-organized hunch, criteria of empirical relevance become important, and the usefulness of a response-class in the systematic study of newborn behavior

comes to be determined by factual considerations. Can the response be observed with high intersubjective agreement? If it is held to indicate individual differences, are values of the response stable for a neonate from occasion to occasion? Is the response related in a systematic way to age, antecedent conditions, or other responses?

The problem of classification posed by Weiss goes beyond the statement of "resemblances," both because the resemblances are not always obvious and because the resemblance selected may not meet empirical requirements for a meaningful response category. The selection of response-classes in the study of newborn behavior will be at first directed by hunch or theory, and those response-classes which are shown to be reliable, stable, and systematically related to other events will survive as useful parts of a system of newborn behavior.

The research to be reported in part here grew from an interest in the early history of tension and tension-control. As Wolf has pointed out, the history of the child can be seen as the history of his responses to developmental "problems," and "the earliest behavioral manifestations of problem-solving are the baby's modes of tension reduction" (8, p. 507). A closely related concern with tension or disequilibrium as important in the understanding of behavior is reflected in the notions of "drive" in contemporary learning theory (e.g., 25) and of "instinct" in psychoanalytic theory (4). In the apparent chaos of neonatal behavior, two primitive generalizations related to these theoretical notions seem to be warranted. First, there are changes from occasion to occasion in the tension or disturbance or disequilibrium of the human newborn. He may be quietly asleep and the next moment awake and screaming; he may be slowly moving his arms and legs and then suddenly startle; he may produce sighs of contentment in the observer or twitches of distress. The second generalization about which we can be relatively certain is that these early variations in tension come to be more and more closely related to environmental events. In fact, to the degree that one is able to set up a stable relation between some aspect of the environment and changes in the infant's level of tension, to that degree will his behavior become meaningful. The infant may become quiet when lifted to the feeding position, for example, or begin to cry "for food" at regular temporal intervals. These basic observations, coupled with the theoretical importance of tracing the early history of tension, led to the first guiding question of the present research: What variables or events influence variation in neonatal tension and tension-control, and what is the early developmental course of this influence? [4]

[4] There are two other closely related ways in which early changes in human behavior can be discussed. The child may be seen as increasingly *in control of his environment*, that is, as becoming more competent. Alternatively, the child may be seen as coming to be increasingly *under the control of his environment*, that is, as becoming more predictable in relation to external events. The differential implications of these two statements of the problem and the "tension" statement will be dealt with in a later paper.

Beyond variations in tension and tension-control in the same children from occasion to occasion, there are theoretical (7, 8) and suggestive empirical (30, 31) grounds for believing that reliable variation in tension occurs between subjects, that is, that tension is a source or index of individual difference. Moreover, in the case of developing tension-control, there is reason to believe that different infants may show different tension-control patterns, some children being more responsive to tactile stimulation, others to gastrointestinal changes, and so on (21). From these considerations emerged the second general question of the research: Can stable individual differences in patterns of tension and tension-control be described for the newborn?

Clearly, these most general statements of the problem do not provide obvious and natural response measures; there remains a wide range of observational protocols which might serve as indicators of tension, and there are perhaps even more variables to be investigated as important in tension-control. For the first empirical analysis of the problem, therefore, the study was narrowed to an investigation of general movement and crying as indicators of tension and of mouthing and hand-mouth contacting as potential tension-control activities. The choice of crying was based on the almost universal popular generalization that it is an expression of distress by the infant; it most assuredly distresses the adult observer. The use of general movement as a partial definition of tension is also justified in common-sense psychology—"a quiet baby is a happy baby"—but it receives support from other sources as well. General movement or activity has been ascribed substantial theoretical importance in an understanding of neonatal functioning; as mentioned earlier, measures of movement have had empirical use of some success; and perhaps of greatest relevance to long-term research goals, we hoped to develop a measure of movement which would be sensitive to changes of short duration in the behavior of the newborn.

The selection of oral behaviors as possibly relevant to tension-control was dictated by both observation and theory. Whatever other disagreements may separate theorists of early development, all agree that mouthing and sucking are fundamentally important responses in the newborn. Moreover, preliminary observation had suggested that two oral responses—hand-mouth contacting and mouthing—could be reliably observed and might reflect systematic individual differences.

Thus, the larger questions raised earlier become reduced to the more readily testable problem: What events or variables control variation in crying, general movement, mouthing, and hand-mouth contacting in the neonate, and how does this control change during the first days of life? There are many routes along which this question can lead, but our initial step was to safeguard more complicated issues by first determining whether the measures chosen meet the empirical criteria stated earlier, that is, by assessing the reliability, stability, and relatedness of our measures. Such is the aim of the present report. Can general movement, crying, and aspects of oral behavior be measured reliably and sensitively? Do these behaviors show

consistent individual differences? Do these behaviors show systematic relations with each other and with situational variables?

Method

Subjects

The subjects of the research were mature, normal infants in the maternity service of Grace - New Haven Community Hospital. Each subject was seen several times during his hospital stay; most subjects were observed at least once each day (*see* below, *Schedule of Observations*). In order to eliminate systematic bias in the selection of subjects, the following criterion was established for inclusion in the sample: that, whenever we began an observation series, the subject chosen be the youngest newborn in the nursery who had not yet been fed. Thus, the age of subjects at the time of their first observation ranged from 1½ hours to 13½ hours, with the median age at 8½ hours. This criterion of selection had the further advantage of permitting the comparison of neonatal behavior before and after the first feeding.

The present report is based on the repeated observation of 50 newborns. The specific observational procedures used were changed as the research became more focused and detailed, with the subjects falling into two main groups: group I ($N = 34$), where our chief interest centered on the observation of hand-mouth contacting; and group II ($N = 16$), where a detailed examination was made of mouthing, crying, contacting, and movement. Each of these groups was subdivided to provide information on special problems (*see* below, *Schedule of Observations*).[5]

Schedule of Observations

Typically, each infant was observed on each of the five days of his hospital stay. Shortly before noon, the infant was brought from nursery to the observation room, which was an unoccupied nursery on the same floor of the hospital. Blankets and other wrappings were removed, and the newborn (dressed in hospital shirt and diaper) was placed supine on the observation pad. If the child was crying, he was held or rocked in arms until he quieted. Once the observation was begun, however, the child was not handled except in cases of acute distress, when a nurse was called to tend to him. The observation period covered from 10 to 30 minutes, depending on the group to which the child was assigned, after which the child was returned to nursery. The schedule of observations for each group of subjects was as follows:

Group Ia ($N = 25$): General observation, followed by five-minute observation (checklist) by at least two observers of hand-mouth contacting. This procedure was repeated, when possible, on days 1, 2, 3, 4, and 5. For 10 sub-

[5] The limitations of the clinical setting in which the research was conducted and occasional apparatus failures led to the discarding of some data on some subjects. As a result, in the analyses presented later, *N*s will vary from those given here.

jects in this group, a pilot run of motion photography (30 seconds in duration) was made on three successive days.

Group Ib ($N = 9$): General observation, followed by five-minute observation (checklist) by at least two observers of hand-mouth contacting. This procedure was repeated, when possible, three times each day (at 8-9 a.m., at 11-12 a.m., and at 5-6 p.m.) on days 1, 2, 3, 4, and 5.

Group IIa ($N = 11$): General observation, followed by five-minute observation (key-depression with record made on an Esterline-Angus operations recorder) of crying, mouthing, and hand-mouth contacting. After this observation, the same measures were taken over five one-minute observation periods, each separated from the next by one minute of no recorded observations. During each of the one-minute observation periods, a 30-second sample of motion photography was obtained. For 10 to 15 seconds of minute-observations 8 and 12, the observer inserted his shielded index finger into the baby's mouth. Figure 1 shows this procedure in diagrammatic form. The procedure was repeated on days 1, 2, 3, 4, and 5.

Group IIb ($N = 5$): Procedure was exactly the same as that for group IIa, except that no oral stimulation was applied and observations were not made during minutes 8, 10, 12.

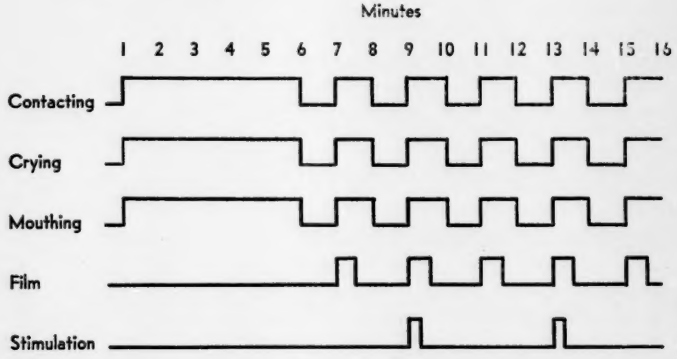

FIGURE 1—Schedule of observational procedures, group IIa.

Response Measures

Hand-mouth contacting. The newborn's response of bringing his fingers or hands to his mouth was recorded in two ways. The first, used with group I, involved placing a checkmark on a mimeographed sheet whenever a contact occurred, recording thereby the finger which made contact and the *frequency* of contacts over a predetermined period of time. For group II, the depression of one of two telegraph keys marked a line on an Esterline-Angus operations recorder (Recorder), indicating which hand made contact, the *frequency* of contacts over a predetermined period, and the *duration* of hand-mouth contacting.

Mouthing. In addition to the movements of the newborn's mouth which can be observed in the feeding situation and which frequently accompany hand-mouth contacting, mouthing movements occur in the newborn in the absence of external stimulation. For group II, a telegraph key was depressed, marking a line on the Recorder whenever the infant's lips moved in a characteristic sucking pattern, either in the presence or absence of external stimulation. In the analysis presented below, the *duration* of mouthing is the measure of this activity.

Crying. For group II, a telegraph key was depressed, marking a line on the Recorder whenever the infant made an audible cry. In order to avoid observer anticipation of crying, the recording key was depressed only on the exhalation part of the crying cycle. As in the case of mouthing, *duration* of crying is the measure used for analysis.

For all duration measures, the Recorder tape was scored by seconds of observation, and a score of 1 was assigned for each second in which hand-mouth contacting, mouthing, or crying was recorded. Thus, for all three duration measures, an infant's score could vary from zero to whatever number of seconds elapsed during the observation. For comparability among observations of different lengths, all scores were reduced to a "per 30-seconds" base. Therefore, for any observation, an infant's crying, mouthing, or hand-mouth contacting duration score varied in a range from zero to 30.

General movement. Each child in group II and 10 children in group I were photographed by a fixed-position overhead camera for one or more runs of 30-seconds duration on several days of the lying-in period. The whir of the camera's spring drive did not, by and large, alter the infant's behavior. No artificial illumination was used other than the overhead lights permanently installed in the nursery. These filmstrips were analyzed to provide a measure of the horizontal displacement of hands and feet over five-second intervals. The details of the technique are described elsewhere (20); although the technique is tedious and time-consuming (requiring the examination and measured comparison of 50 film-frames per child per observation), it provides a reproducible and permanent record of the infant's movements.

Results

Reliability of Measures

Hand-mouth contacting. The reliability of the checklist technique was high; correlations between pairs of observers were invariably above .90, and, for the two observers who did the bulk of observations for group Ia, the interobserver correlation was .96. The Recorder technique used for group II showed similar intersubjective reliability, with the protocols of the two most frequent observers showing a correlation of .96.

Mouthing. The occurrence of mouthing was the event most difficult to discriminate in the study. Nonetheless, test runs for reliability produced an interobserver correlation coefficient of .93.

Crying. Although there is some ambiguity in the observation of crying (e.g., unvoiced cries, "crying face" without sound), observers did not have the reservations typical for the observation of mouthing, and test runs for reliability showed an interobserver correlation of .99.

Movement. The assessment of reliability for the movement measure was made by comparing the movement values assigned to children by two analyzers after they had independently sketched and measured the appropriate frames of film. The discriminations required for this task were matching one picture against the next and measuring linear displacement of limbs with a millimeter rule [*see* (20) for further details]. The correlations which resulted from this assessment ran above .96.

The measures set up for general movement, crying, mouthing, and hand-mouth contacting were shown to be very highly reliable; moreover, the absolute values recorded by different observers did not show more than negligible variation in means and variances.

Individual Stability of Measures

Being assured that reliable protocols can be established for movement, crying, hand-mouth contacting, and mouthing, we put this next question to the data: Are these measures stable from infant to infant over the lying-in period; that is, is there evidence of consistent individual variation among newborns for these response categories?

Hand-mouth contacting. Hand-mouth contacting is by no means an infrequent response in the human newborn. Without exception, every infant seen in pretest series and in the research proper was observed to bring his hand to his mouth at least once during the five-minute observation periods. Moreover, the variation in frequency of the response is great; among the 32 subjects who were observed on each of the five days of hospital stay, the average number of hand-mouth contacts during five minutes ranged from 3 to above 30, with two subjects making contact 60 times in a single observation period.

In group I (checklist group), 16 newborns were seen on all five days at the midday observation period; this provides an array of 80 contacting scores which can be evaluated in two ways to indicate individual stability. The first of these is a rank order procedure, in which frequency of hand-mouth contacting is ranked for each day, and the coefficient of concordance (18) computed for the resulting array of ranks. This statistic had a value of .51 ($p < .01$), indicating that a significant proportion of variation in ranked frequency of hand-mouth contacting could be assigned to individual differences. The same finding can be shown in an analysis of variance of the day-by-subject array of hand-mouth contacts; such an analysis shows a significant variance ($F = 3.47$, $df = 15/60$, $p < .01$) ascribable to between subject variation.

Group II (Recorder) provided an opportunity to check the replicability of this finding. Sixteen subjects were seen on five days of hospital stay, and

the resulting coefficient of concordance was once again significantly different from chance ($W = .34$, $p < .05$). The analysis of variance of raw contacting scores showed a parallel outcome ($F = 2.19$, $df = 15/60$, $p < .05$).

In evaluating these findings, it must be remembered that they are based on a relatively brief sample of the infant's total stay in the hospital (about 0.5 per cent), and under circumstances (e.g., variation in maternal handling) which introduce considerable unsystematic variation. Therefore, the stability demonstrated here is almost certainly an underestimate of the stability to be expected from a longer time sample or in an artificially controlled setting.

It is interesting to note that the data from group II provided no evidence of significant individual stability for *total duration* of hand-mouth contacting ($W = .22$) nor for *mean duration per contact* ($W = .10$). This result suggests that getting the hand to the mouth may be under the control of variables different from those which control the maintenance of a contact once made.

Mouthing. Some activity of the mouth, whether responsive to external stimulation or not, is a strongly prepotent response in the human newborn. Among the 16 infants in group II, the five-day average duration of mouthing during a five-minute observation ranged from 56.4 seconds (18.8 per cent of time observed) to 189.6 seconds (63.2 per cent) with a mean at 112.5 seconds (37.5 per cent). For this measure, a coefficient of concordance computed on an array of ranks does not reach statistical significance ($W = .21$); the corresponding analysis of variance shows a marginally significant variance attributable to between subject variation ($F = 2.09$, $df = 15/60$, $p < .05$). For the conditions of observations obtaining in the present study, mouthing is apparently not as sensitive a measure of stable individual variation as is hand-mouth contacting.

Crying. For the 16 subjects in group II, the crying measure showed wide variation, with mean duration per five-minute observation over the lying-in period ranging from 6.8 seconds (2.3 per cent of time observed) to 117.6 seconds (39.2 per cent) with a mean at 54.5 seconds (18.2 per cent). The distribution of crying durations was irregular in shape, with 27 of 80 observations showing no crying at all. No significant individual differences were found on the crying measure. General observation of the infants suggested that this response is both too infrequent and too much under the control of fluctuating external stimulation (e.g., wet diapers) to be a convincing index of individual variation.

Movement. The 30-second filmstrips run for each subject in group II on each day of his hospital stay provided a measure of neonatal movement. The sum of displacements over five-second intervals of right and left palm and right and left ankle was the primary score submitted to analysis. For 10 subjects in group IIa, this score was computed for five daily observations and these scores combined to provide a single daily value of movement for each infant. The resulting coefficient of concordance was .53 ($p < .01$), con-

firming the general observation that a significant portion of the variation in newborn movement can be assigned to individual variation. The corresponding analysis of variance shows the same result ($F = 3.86$, $df = 9/36$, $p < .01$). It should be noted, too, that this score combines observations of "spontaneous" movement with observations of movement following external stimulation (i.e., insertion of the observer's shielded finger in the infant's mouth at observation minutes 8 and 12 each day). Individual stabilities were clearly potent enough to override the effect of this intervention, and a high degree of confidence can be placed on the finding of consistent subject differences.

This stability in neonatal movement can be seen even more dramatically by examining the scores of groups IIa and IIb for the *first* 30-second observation each day. The coefficient of concordance for 15 subjects was .49 ($p < .01$); the analysis of variance showed significant variation between subjects ($F = 3.33$, $df = 14/56$, $p < .01$). Thus, a stable measure of individual variation in newborn movement can be computed, *based on a 30-second sample* on each day of hospital stay.

The range of variation among neonates in movement over a five-second measured interval, averaged for the lying-in period (groups IIa and IIb, first filmed observation each day) was from 1.50 cm./limb to 5.70 cm./limb, with a mean at 3.65 cm./limb.

In assessing the four measures under examination in the present study for their sensitivity in showing stable individual variation, we may conclude that hand-mouth contacting and general movement vary in a significant way from child to child, that mouthing is a marginally sensitive measure of individual variation, and that crying as measured in the present setting does not provide a reliable index of individual variation.

Relatedness of Measures

Day-to-day effects. During the first five postnatal days, the newborn does not merely age; he is trained to a new feeding technique, he is extensively handled, if he is a male he undergoes minor surgery, and so on. It is therefore difficult to assign sources of variation in day-to-day effects. However, several meaningful questions may be asked. Does the behavior show regular and systematic change over days? Are there significant differences between day 1 (when, for our sample, no feeding had occurred) and later days? Do relationships among variables change over time? There are 15 subjects in group II on whom complete data are available which can be used to answer these questions.

Figure 2 shows changes over days in mouthing, hand-mouth contacting, and crying. As might be anticipated from the earlier finding of great variability of mouthing and crying, neither of these measures shows significant day effects. Whether the increase in crying scores between day 1 and day 4 which appears to visual inspection can be brought to significance by a larger sample, whether of subjects or time, remains to be investigated. Of more

interest is the finding that mouthing is insignificantly different from day 1 to later days. It might have been anticipated that feeding experience would affect mouthing, either increasing it as a function of the reinforcement operation of feeding or decreasing it as a function of the increasing frustration of nonnutritive mouthing. Neither effect is evident.

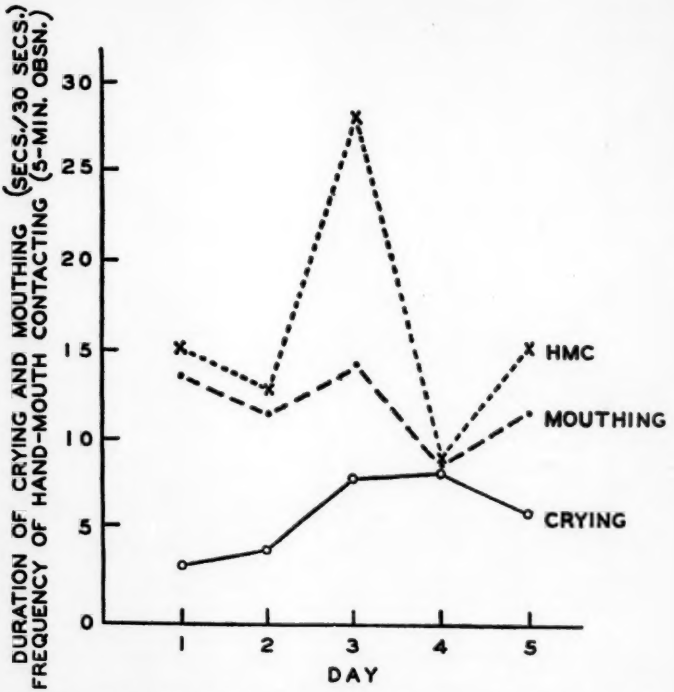

FIGURE 2—Changes in hand-mouth contacting, mouthing, and crying over the first five days of life.

The hand-mouth contacting measure is somewhat more difficult to interpret. The differences shown in Figure 2 are significant ($F = 4.51$, $df = 14/56$, $p < .01$), although it is apparent that the sharp increase in contacting on day 3 accounts for most of the variation. An examination of other responses and of hospital routines has shown no covariation which would readily account for this surprising hump. Nonetheless, it is interesting that, as in the case of mouthing, the increasing number of feeding "trials" the child has does not seem to influence hand-mouth contacting in any

systematic way. Contacts on day 1 are as frequent as those on day 2, day 4, or day 5. This finding is important in that it casts serious doubt on an assumption that hand-mouth contacting is "secondary" to feeding or learned in the feeding situation.

The day-to-day plot of movement for group IIa subjects shown in Figure 3 is relatively free of ambiguity. There is a systematic and significant ($F =$

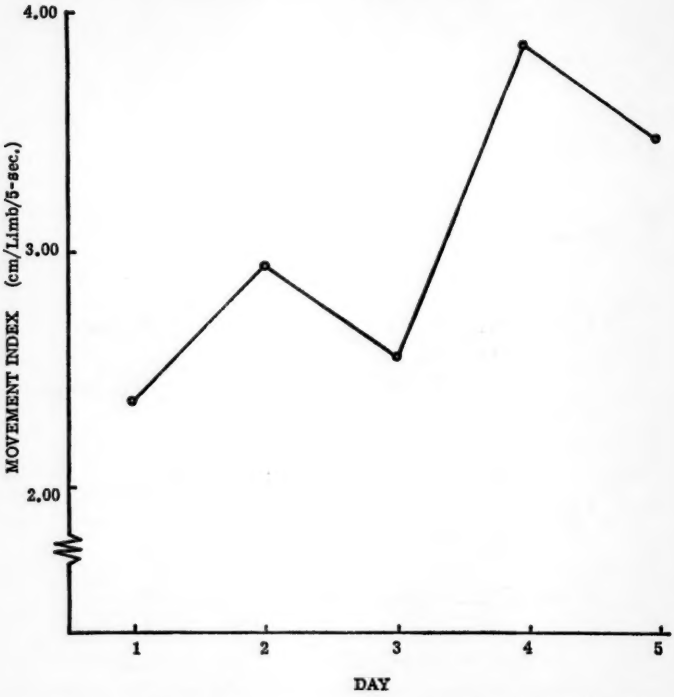

FIGURE 3—Changes in neonatal movement over the first five days of life.

3.03, $df = 4/36$, $p < .05$) increase in neonatal movement during the early days of postnatal life. This finding, beyond confirming the observations of other investigators, adds to our confidence that the movement measure used in the present research is a sound measure of individual and situational variation.

Interrelations among measures. Do the response measures hang together in a way which would suggest a primitive "typology" of newborns? For each child in group II ($N = 15$), a single index was computed of crying,

mouthing, contacting, and movement by summing daily values over the entire hospital stay. The rank correlations (rho) among these *over-all* measures were as follows:

Contacting-Mouthing	.36	Mouthing-Movement	.49
Contacting-Crying	.19	Movement-Crying	.18
Contacting-Movement	.19	Mouthing-Crying	—.27

Although none of these correlations reaches conventional levels of statistical significance, there is a suggestion that the more active infants mouth more and that the infants who mouth more often make contact more often. There is, furthermore, no strong evidence that a simple "general factor" will account for the relations suggested. It is interesting to note that the low correlation between movement and contacting argues against a random-hit interpretation of hand-mouth exchange; it does not seem to be the case that an infant who moves a lot happens to strike his mouth frequently.

The complexity of interrelations among newborn responses is seen in an examination of day-to-day changes in correlations (Table 1). For the comparisons (again, the coefficient is rho) crying was not included because of the relatively high incidence of zero scores. The major implication of this table is clear; a tendency for the three measures to be related fairly closely early in the postnatal period dissipates toward chance by the fifth day of life. Two hypotheses can be advanced to account for this striking shift. The interrelation at birth and shortly thereafter may be under the control of some transiently effective variable (e.g., maternal anesthetic), or the environmental contingencies of the first five days may be such as to disturb a congenitally organized response-set. The decay in interrelations among measures cannot be assigned to increasing unreliability of measures. For both movement and frequency of contacting, the *individual* stabilities between day 4 and day 5 are as great as or greater than the stability among earlier days. An accurate interpretation of these findings must await further empirical examination of newborn behavior and its environment.

TABLE 1

RANK-ORDER CORRELATION COEFFICIENTS (*RHO*) FROM DAY TO DAY AMONG THREE MEASURES OF NEWBORN BEHAVIOR

	Movement-Contacting	*Movement-Mouthing*	*Contacting-Mouthing*
Day 1	.57*	.55*	.68**
Day 2	.25	.35	.59*
Day 3	—.09	—.21	.70**
Day 4	—.16	.17	.32
Day 5	.11	—.03	.18

NOTE.—One asterisk indicates significance at .05 level, two asterisks significance at .01 level. $N = 16$ for days 1, 2, 3, and 4; $N = 13$ for day 5.

The Effects of Stimulation

On two observations each day, the infants in group IIa were given an opportunity to suck on the observer's shielded index finger for 10 to 15 seconds. Under the conditions of the present study, no stable immediate effect on neonatal movement could be detected; although the infants seemed to quiet somewhat when the stimulation was present, this general observation was not confirmed by an examination of the filmstrips.

Sex differences. Some variation between the behavior of males and the behavior of females was noted in each group but no sex differences were found to be both statistically significant and replicated.

Discussion

Two of the measures of newborn behavior under study, movement and hand-mouth contacting, met reasonable demands for reliability of observation and for sensitivity to individual variation. The other measures, mouthing and crying, can be observed in the neonate with satisfactory precision, but did not show in the setting described here significant sensitivity to individual differences. Of the four, movement showed systematic changes over the lying-in period.

The finding of stable individual differences in movement and hand-mouth contacting among human newborns confirms the often stated but infrequently demonstrated proposition that human beings show substantial variations one from another aboriginally. Even for the relatively homogeneous group of children observed for this research, wide and stable differences were shown. Knowing this, it is important to turn to two further questions. Do these early differences make any difference later? Does it make good sense to consider these variations, whether from time to time or from baby to baby, as representing differences in tensions? The first of these questions requires longitudinal study; the second demands an examination in some detail of the relation between activity and contacting on one hand and environmental events and other neonatal responses on the other. If neonatal movement is to serve as an index or partial interpretation of the term "tension," then it should be possible to demonstrate systematic relations to privation, to noxious external stimulation, and to the environmental operations used by parents and clinical practitioners to reduce distress in infants. On the basis of the present study, we can proceed to an attack on these next empirical questions with some confidence in the precision and the stability of measures for hand-mouth contacting and infant movement.

Summary

An initial attack on the problems of newborn behavior was made in an examination of the reliability, individual stability, and relatedness of four response measures, movement, crying, mouthing, and hand-mouth contact-

ing. Infants were seen repeatedly during the lying-in period. All measures showed very high intersubjective reliability. The measures for movement and hand-mouth contacting were sensitive to individual variation over the first five days of life; the measures for crying and mouthing were not. The movement measure showed systematic change over the lying-in period; for the other measures, neither a regular increase nor a regular decrease in occurrence was found, putting in some doubt the suggestion that mouthing and hand-mouth contacting are phenomena secondary to feeding.

The relation of the findings to the problem of tension and tension-control in the newborn was discussed.

References

1. Balint, M. Individual differences of behavior in early infancy and an objective method for recording them. *J. genet. Psychol.*, 1948, 73, 57-59; 81-117.
2. Dennis, W. A description and classification of the responses of the newborn infant. *Psychol. Bull.*, 1934, 31, 5-22.
3. Freud, S. *Three essays on the theory of sexuality.* London: Allen & Unwin, 1905 [1949].
4. Freud, S. *An outline of psychoanalysis.* Norton, 1949.
5. Fries, M. E. Psychosomatic relationships between mother and infant. *Psychosom. Med.*, 1944, 6, 159-162.
6. Fries, M. E. Diagnosing the child's adjustment through age-level tests. *Psychoanal. Rev.*, 1947, 34, 1-31.
7. Fries, M. E., & Lewi, B. Interrelated factors in development. *Amer. J. Orthopsychiat.*, 1938, 8, 726-752.
8. Gardner, G. E. Problems of early infancy. *J. Amer. Psychoanal. Ass.*, 1955, 3, 506-514.
9. Gesell, A. *The embryology of behavior.* Harper, 1945.
10. Gesell, A., & Halverson, H. M. The daily maturation of infant behavior. *J. genet. Psychol.*, 1942, 61, 3-32.
11. Graham, F. K., Matarazzo, R. G., & Caldwell, B. M. Behavioral differences between normal and traumatized newborns. *Psychol. Monogr.*, 1956, 70, No. 20 and No. 21.
12. Greenacre, P. The biologic economy of birth. *Psychoanal. Study Child*, 1945, 1, 31-52.
13. Halverson, H. M. Studies of the grasping responses of early infancy. *J. genet. Psychol.*, 1937, 51, 371-449.
14. Hoffer, W. Mouth, hand and ego-integration. *Psychoanal. Study Child*, 1949, 3/4, 49-56.

15. Irwin, O. C. The amount and nature of activities of new-born infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monogr.*, 1930, 8, 1-92.

16. Irwin, O. C. The distribution of the amount of motility in young infants between two nursing periods. *J. comp. Psychol.*, 1932, 14, 429-445.

17. Jensen, K. Differential reactions to taste and temperature stimuli in newborn infants. *Genet. Psychol. Monogr.*, 1932, 12, 363-479.

18. Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.

19. Kendler, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.*, 1952, 59, 269-277.

20. Kessen, W., Hendry, A. L. S., & Leutzendorff, A. M. Measurement of movement in the human newborn: a new technique. *Child Develpm.*, 1961, 32, 95-105.

21. Leitch, M., & Escalona, S. K. The reaction of infants to stress. *Psychoanal. Study Child,* 1949, 3/4, 121-140.

22. Logan, F. A. *Incentive.* Yale Univer. Press, 1960.

23. Lustman, S. L. Rudiments of the ego. *Psychoanal. Study Child,* 1956, 11, 89-98.

24. Marquis, D. P. Can conditioned responses be established in the newborn infant? *J. genet. Psychol.*, 1931, 39, 479-492.

25. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* Wiley, 1951. Pp. 435-472.

26. Piaget, J. *The origins of intelligence in children.* International Universities Press, 1952.

27. Pratt, K. C., Nelson, A. K., & Sun, K. H. The behavior of the newborn infant. *Ohio State Univer. Stud., Contr. Psychol.*, 1930, No. 10.

28. Spitz, R. A., & Wolf, K. M. Autoerotism. *Psychoanal. Study Child,* 1949, 3/4, 85-120.

29. Weiss, A. P. The measurement of infant behavior. *Psychol. Rev.*, 1929, 36, 453-471.

30. Wolf, K. M. *The origins of individuality.* Mimeo., 1957.

31. Wolff, P. H. Observations on newborn infants. *Psychosom. Med.*, 1959, 21, 110-118.

FORM, COLOR, AND SIZE IN CHILDREN'S CONCEPTUAL BEHAVIOR [1]

JEROME KAGAN* *and* JUDITH LEMKIN
Fels Research Institute

Although conceptual behavior of adult humans has always been a frequent problem for investigation, studies of conceptual processes in the child have attracted considerably less attention. The work of Piaget (4) has recently catalyzed a vigor in this area and has made psychologists aware of the lack of reliable empirical information available on children's conceptual behavior. This very brief note is the first in a series of studies which seeks to understand the way in which the child conceptualizes his environment. This particular research focuses on the tendency to use form, color, or size in classifying stimuli into conceptually similar groups.

Over 30 years ago, Brian and Goodenough (1) found that the child under 3 years of age preferred form to color in organizing his world into groupings of "similar" stimuli. From 3 to 6 years of age, the child shifted to color over form as the preferred basis for categorization. After 6, form once again became dominant over color. The recent interest in conceptual behavior is flavored with a concern for individual differences rather than statements about children in general. We are now more suspicious of statements which purport to hold for all children and expect that there will be important sex, age, and even personality differences in conceptual behavior. Although Brian and Goodenough reported no significant sex differences, a recent investigation (3) with children ages 7 to 11 found that more girls than boys used form rather than color as a basis for classifying stimuli into conceptual classes. The present study is, in part, a replication of this experiment with a younger group of children.

METHOD

The subjects (*Ss*) were 35 girls and 34 boys ranging in age from 3-9 to 8-6 with a median age of 5-8. All *Ss* were enrolled in either a nursery or the first grade of an elementary school. The junior author was the experimenter (*E*), and she initially spent several hours in the classes of the *Ss*

* Fels Research Institute, Yellow Springs, Ohio.

[1] This research was supported, in part, by research grant (M-1260) from the National Institute of Mental Health, United States Public Health Service.

Child Develpm., 1961, 32, 25-28.

in order to promote familiarity and facilitate rapport for the individual testing situation which was to follow. Each *S* was tested individually in a room provided by the school. The *E* spent a few minutes gaining the child's confidence and then initiated the concept formation task.

Each of the nine stimuli in the series consisted of pieces of white paper upon which were pasted paper cutouts of geometrical forms differing in color, shape, and size.[2] The first three stimuli had two comparison forms on the top of the paper and a standard stimulus at the bottom. Stimuli 4 through 9 had three comparison forms on the top of the paper and the one standard stimulus at the bottom. The method used here was essentially similar to the procedure of Brian and Goodenough. For stimuli 1 through 3 the instructions were: "Here are some cutouts. Do you see this cutout down here? Which one up here (*E* points to the top of the page) is the same as the one down here?" For items 4 through 9 the instructions were: "Now we have one cutout down here and three up here. Which one up here is the same as the one down here?"

Each of the nine stimuli contained forms which could be classified with the standard stimuli on the basis of form, color, or size. For each stimulus only two of these three alternatives were possible as correct solutions. For stimulus 1, color was opposed to form; for stimulus 2, color vs. size; stimulus 3, size vs. form; stimulus 4, color vs. form; stimulus 5, form vs. size; stimulus 6, color vs. form; stimulus 7, form vs. size; stimulus 8, color vs. size; stimulus 9, color vs. size. The *E* recorded the child's answer and asked why the child had chosen that stimulus. All the children understood the task and performed appropriately, but usually were unable to answer why they had made the response they did.

Results

Table 1 presents the percentage of boys and girls who chose form, color, or size for each of the nine stimuli. All children strongly preferred form to either color or size, and size was rarely used as a basis for conceptual similarity. For any one stimulus, there were no statistically significant sex or age differences in choice of color, form, or size. However, when the total number of form, color, and size responses were summed for all nine stimuli, two interesting results emerged.

Table 2 presents the average number of form, color, and size choices for boys and girls who were divided at the median into young and old groups. There was no difference between the responses of young vs. old boys. However, older girls were much less likely than younger girls to choose color ($p < .02$; two tails). Secondly, older boys were much more likely than older girls to use color as the basis for conceptualization ($p < .05$; two tails).

[2] The standard stimuli sampled one of three forms (triangle, square, or circle), one of four colors (orange, green, red or blue) and one of two sizes (large or small). The area of the large forms was approximately six times the area of the smaller forms.

TABLE 1

PERCENTAGE OF BOYS AND GIRLS CHOOSING FORM, COLOR, OR SIZE AS BASIS OF CLASSIFICATION

Stimulus	*Alternative*	*Preferred Choices*	*Boys*	*Girls*
1	F vs. C	F	64.7	85.7
2	C vs. S	C	73.6	60.0
3	F vs. S	F	73.6	91.4
4	F vs. C	F	64.7	80.0
5	F vs. S	F	67.7	80.0
6	F vs. C	F	58.8	77.3
7	F vs. S	F	73.6	88.6
8	C vs. S	C	53.0	51.3
9	C vs. S	C	64.7	68.5

This finding agrees with the results of Honkavaara who found that more girls than boys (aged 7 to 11) preferred form to color as a basis for conceptualization. The fact that color is not a preferred hypothesis for children this age is supported by the unexpected finding that, when color and size were present as the only bases for similarity, over 20 per cent of the *S*s said they did not know what the answer was and could not offer an answer to *E*. That is, when the stimuli did not allow for the use of form but did allow for the use of color, one fifth of the children did not know how to respond. This indicates that a color hypothesis is not merely weaker than form but, for some *S*s, is unavailable when the child is asked the question, "How are these stimuli the same?"

TABLE 2

MEAN NUMBER OF FORM, COLOR, AND SIZE CHOICES FOR ALL NINE STIMULI BY SEX AND AGE

	Boys		Girls	
	Young	*Old*	*Young*	*Old*
Form	4.2	4.3	4.3	5.7
Color	3.4	3.6	3.4	1.6
Size	0.4	0.9	0.6	1.2

Discussion

The results indicate that the preschool child's understanding of the phrase "same as," in the present context, is influenced primarily by the shape or form of the stimulus rather than its color or size.

The sex differences between older boys and girls is difficult to explain. One post hoc argument is presented. Since the verbal skills of young girls proceed precociously in relation to boys, it is possible that the girls implicitly apply the language labels, square, triangle, circle, to the stimuli more often than boys. Thus, for the girls, the stimuli are more likely to derive their meaning from the label attached to them rather than through the more direct physical quality of color. This explanation agrees with Sigel's (5) finding that, when school age children were asked to sort diverse stimuli into conceptually similar groups, the semantic meaning attached to the stimuli crowded out other bases for conceptualization (color, type of material, texture). Sigel concluded that the label attached to the stimulus is the primary basis for conceptualization in school age children. Brian and Goodenough reported that the brighter children were more likely to prefer form to color, and this incidental datum supports the present post hoc argument.

Summary

This paper describes an experiment on the conceptual preferences of children when they were asked to group geometric stimuli differing in form, color, or size into similar pairs. Subjects, 35 girls and 34 boys (ages 3-9 to 8-6), were presented with nine stimuli in which a standard stimulus was presented simultaneously with other stimuli differing in form, color, and size. The child was asked to select the comparison stimulus which was the "same as" the standard. The result indicated that for both boys and girls form was distinctively preferred to color as a basis for similarity and that color was preferred over size. For boys, there was no age difference in this response pattern. However, older girls were less likely than younger girls to use color as a basis for conceptualization. Older boys, moreover, were more likely to use color than were the older girls. These results agreed with a recent investigation in this area, and it was suggested that the sex difference was a function of the girls' implicit labeling of the forms as square, triangle, or circle, which presumably facilitated the use of form as a basis for similarity.

References

1. Brian, C. R., & Goodenough, F. L. The relative potency of color and form perception at various ages. *J. exp. Psychol.*, 1929, 12, 197-213.
2. Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* Wiley, 1956.
3. Honkavaara, S. A critical reevaluation of the color and form reaction, and disproving of the hypotheses connected with it. *J. Psychol.*, 1958, 45, 25-36.
4. Piaget, J. *The construction of reality in the child.* Basic Books, 1954.
5. Sigel, I. E. The dominance of meaning. *J. genet. Psychol.*, 1954, 5, 201-207.

DEVELOPMENTAL CHANGES IN PREFERENCE FOR GOALS DIFFICULT TO ATTAIN[1]

DONALD C. SMITH* *and* LUCY WING**
Syracuse University

In the 1930's Wright (3) conducted a series of experiments with children and adults to test the hypothesis that in a choice situation an obstructed goal has a stronger positive valence than an accessible one. His findings, which generally supported the hypothesis, also indicated that preference for a difficult goal is positively associated with increasing age. Adults in his study more consistently chose the difficult goal than did children. Preferences were also found to vary with the degree of difficulty of the barrier, with the nature of the goal object (*a goal in itself* or *a means to an end*), and with manner of expression of choice (verbally or through action).

Child and Adelsheim (2) replicated Wright's experiments with slight variations. They found that nursery school children consistently preferred the nearer, more available goal. A later developmental study by Child (1) indicated that preference for the nearer of two goals was much more frequent among younger children than among older children.

The present experiment is an extension of Child's developmental study. One of the acknowledged limitations of his research was the use of only one type of barrier. The present research employed a barrier of difficulty

* Bureau of Educational Research and Service, Ohio State University, Columbus 10.

** Department of Psychology, Oswego State Teachers College, Oswego, New York.

[1] The writers are greatly indebted to Dr. George G. Thompson for valuable advice and criticism. They also wish to thank Principal Hilda Jacobson and Principal John Dunn of the Liverpool School System for their cooperation in making available subjects for the study.

Child Develpm., 1961, 32, 29-36.

as well as a barrier of distance. Goal objects were employed which were classified as goals in themselves and as means to ends. An attempt was also made to control more rigorously the following variables which previous experiments had found related to preferences for more difficult goals: sex, intelligence, socioeconomic status, racial origin, and academic achievement of subjects, and sex of the experimenters. As additional controls, an equal number of goal and means objects were included, barriers were selected which were well within the capacity of the youngest child, and communication was prevented between subjects prior to testing.

The main purpose of the present experiment was to investigate developmental changes in preference for goals which are more difficult to attain. The experimental hypothesis was that preference for difficult goals increases with increasing chronological age. It was assumed that, through experiences of social approval, children in our society learn to place a high value on the achievement of difficult things. As they grow older, and there are more occasions for social reinforcement of their preferences, the frequency of choice of difficult goals was expected to increase.

Three other questions were considered in the analysis of preferences for difficult goals: (a) the effect of the sex of the examiner, (b) the effect of the nature of the task, (c) the effect of goal objects which are ends in themselves or means to an end.

Subjects

The subjects were 80 boys enrolled in the Liverpool, New York Public Schools. Groups of 20 boys from grades 2, 4, 6, and 8 were closely equated on intelligence and socioeconomic status. Intelligence test scores from the California Test of Mental Maturity were used. The mean intelligence quotients (IQs) for grades 2, 4, 6, and 8 were 118, 114, 114, and 115, respectively. IQs for the total group ranged from 94 to 144. Socioeconomic status was based on the occupations of fathers of *Ss*, rated according to the University of Chicago revision of the Warner, Meeker, and Eells Scale for Rating Occupations (4). Occupational ratings of fathers ranged from 2 through 5 on the Index of Status Characteristics, and the mean occupational ratings were identical for the four grade levels. The average socioeconomic ratings fell in the lower middle-class range.

Only white, native-born *Ss* were included and only those whose school achievement was within the average to above-average range on the basis of standardized achievement tests and judgments of teachers. Any child who in the opinion of the teacher, guidance counselor, or the helping teacher was physically handicapped or seriously emotionally disturbed was excluded from the study. The age range for the total group was 7-7 to 14-4 years. Mean ages for grades 2 through 8 were 8-1, 10-0, 11-8, and 13-9, respectively. Standard deviations in months for the four grade levels were 3.0, 4.1, 6.6, and 3.2.

Experimental Procedures

Subjects were randomly assigned to the male and female experimenters so that 10 boys at each grade level were seen by each. Boys were taken from their classrooms one at a time and followed each other in rapid succession. While one subject was being seen, another was taken to the experimental room. Schedules were carefully checked so that playground, lunch, or other free periods did not intervene before all subjects in one classroom had completed the experiment; therefore, subjects already taken from the home room had no opportunity to discuss the situation with future subjects.

After personal data had been secured, the experimenter explained to each subject that the purpose of the study was to find out the kinds of things boys his age liked to do and emphasized that it had nothing to do with school work. Previously, teachers had been instructed to underplay the situation, and no undue curiosity was aroused. Most of the children were accustomed to occasional withdrawal from their classrooms for individual testing.

The four experimental tasks placed each subject in the position of having to choose between two goals, one accessible and the other relatively obstructed. The obstructed goals involved the barriers of greater distance or greater difficulty. The Table and Ladder situations, which represented the barrier of distance, were identical to the experimental tasks used by Child (1). On the Formboard and Assembly, the barrier was considered to be the number of puzzle pieces available for completion of the tasks. For the former two situations, candy served as the goal object; for the latter two situations, the optional number of puzzle pieces were considered to be *means to an end* goal objects.

Table Situation. *S* was asked to stand at one end of a table. A piece of chocolate-covered cream candy was placed directly in front of him and an identical piece of candy was located at the center of the opposite end of the table. The boy could reach the nearer candy easily but had to walk around the table to secure the more distant piece. Both pieces of candy were wrapped in cellophane so there would be no rejection for sanitary reasons. *E* said, "Now would you please stand here at the end of the table? Do you see the two pieces of candy? You may have either one you want." If *S* asked if the two pieces were alike, a noncommittal reply was given, "Well, I don't know. They look about the same."

Ladder Situation. Next, *S* was asked to stand by a three-step ladder which was in front of a book shelf. One piece of chocolate candy was placed within easy reach on a low shelf. The other candy was put on a high shelf so that *S* had to climb the steps to secure the more distant goal. The placement of the candy on the shelves was adjusted to the height of subjects so that the distance would be the same for both younger and older children. *E* said, "Now, come and stand over here" (indicating a place directly in front of the shelves). "Do you see those two pieces of candy? You could get one by climbing the steps. You could get this one by reaching right over here. You may take either one you want."

Formboard Situation. The equipment was a nine-inch formboard, and subjects had the choice of completing this with either eight or 16 triangular pieces. The pieces were stacked in two separate piles; whether the pile with the most pieces was to the right or left of *S* was predetermined by random assignment. *E* said, "Now, let's sit down here at the table. Here is a puzzle we would like you to try. The pieces in both piles will fit into this space. This pile has more pieces than the other one. I'd like you to finish the puzzle by using the pieces in one of these piles. You may choose either one of the piles you wish. Which one do you want to use? (Pause) All right, put the puzzle together." *E* did not show interest in a subject's proficiency in doing the puzzle. If a child had much difficulty and showed any signs of frustration, *E* helped him to complete the formboard.

Assembly Situation. This task involved a simple serial-order assembly consisting of wooden forms which fitted into a grooved base. The subject could select an assembly with 10 or 15 pieces. The size and shape of these pieces were identical in both cases. The following instructions were given: "Here's another kind of puzzle. What you do is take all the pieces out and then put them back together. This one has more pieces in it than the other one. You may choose either one. Which one do you want to do? (Pause) All right. Go ahead and do it." The 10- and 15-piece assemblies were randomly assigned to the right or left side of subjects.

Results

The main purpose of the experiment was to establish whether choices of more difficult goals increased proportionally with age. This first hypothesis was not supported, as shown in Figure 1. The mean percentage of children choosing the more difficult goals increased from 32 per cent at

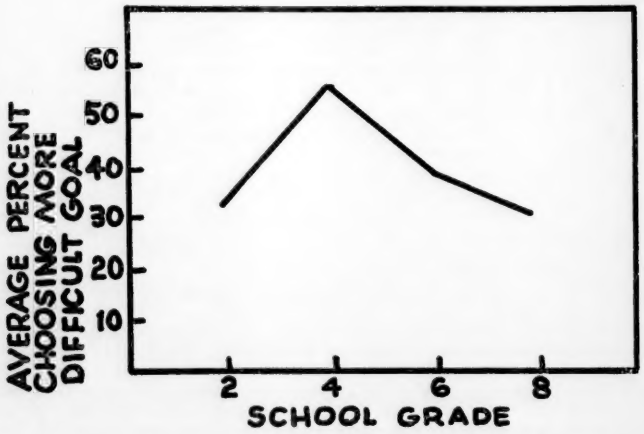

Figure 1—Changes with school grade in choice of more difficult goal.

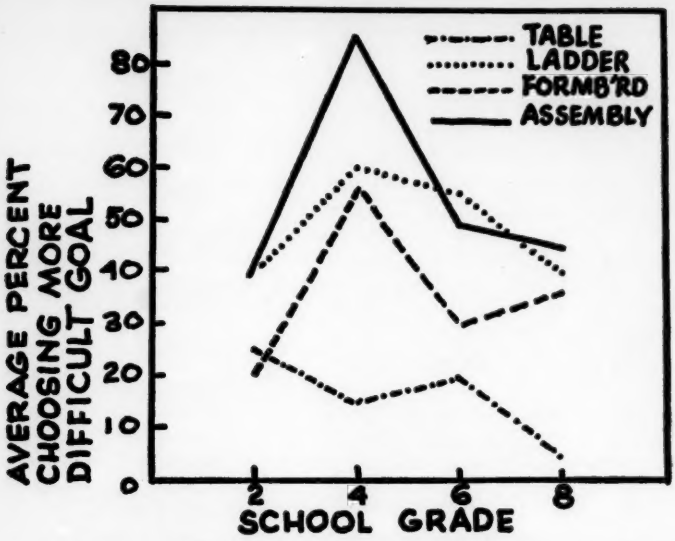

FIGURE 2—Changes with school grade in choice of more difficult goal in the four different situations.

grade 2 to 55 per cent at grade 4 but decreased to 39 per cent and 31 per cent at grades 6 and 8. Thus, when responses to all four tasks are combined, the percentage of boys choosing the more difficult goal was in the direction predicted for grades 2 and 4 but decreased progressively for grades 6 and 8.

As may be seen in Figure 2, slight variations occurred in the cases of the Table situation and the Formboard task. The marked increase at the fourth-grade level was found with all the experimental tasks except for the Table situation. Although the general trend is downward from the fourth grade on, slight deviations may also be noted on the Formboard and Table situations.

Child (1) reported that the sex of the experimenter was a variable which had an important bearing on the responses of subjects. His experiment indicated that subjects were more likely to choose the more difficult goal with a female than with a male experimenter. However, in the present study, no significant differences were found between the responses of the subjects who worked with the male experimenter and those who worked with the female experimenter.

The second experimental variable investigated was the influence of the nature of the task upon preference for more difficult goals. It was suspected that preference for the more difficult goal would vary considerably in terms

of the type of task which confronted the child. Chi square values for the Table and Assembly tasks were not statistically significant. However, on two of the tasks, Table and Formboard, fewer subjects selected the difficult goal than would have occurred by chance alone ($p < .01$). Thus, the nature of the task appeared to be definitely associated with the preference for specific goals.

Wright (3) concluded that difficult goals are selected more frequently when the goal object is an end in itself than when it is a means to an end. In order to obtain additional data related to this hypothesis, two "end" goal objects and two "means" goal objects were included in the present experiment. The Table and Ladder situations involved a choice between two goal objects which were *ends* in themselves. The Formboard and Assembly tasks involved a choice between two goal objects which were *means* to the completion of the task.

For the total sample, 32 per cent of the subjects chose the more difficult goal when the goal was an end in itself. When the goal object was a means to an end, 45 per cent chose the more difficult goal. Almost all subjects consistently preferred the difficult goal more frequently with the *means* goal objects than with the *ends* goal objects, except for a slight negative difference for subjects in grade 2. The critical ratio test was applied for differences in correlated proportions, and the differences were found to be statistically significant ($p < .05$). The data, therefore, are contradictory to Wright's hypothesis. It could be legitimately argued, however, that the greater preference for the difficult goals on the Formboard and Assembly task was due to the difference in types of barriers rather than to the difference in goal objects.

Discussion

Despite differences in sampling and in experimental tasks, the present findings are in close agreement with those obtained by Child (1). He found that about 20 per cent of first and second graders chose the more distant or more difficult goal. In grades 3 and 4 this proportion increased to slightly above 50 per cent. After this maximum had been reached, the proportion tended to fluctuate irregularly in succeeding years. Thus, in Child's study, the preference for the more difficult goals increased to a point somewhere above eight or nine years and fluctuated irregularly thereafter. In the present experiment, preference for the more "distant" or more "difficult" goals increased to a maximum of 55 per cent at grade 4 (approximately age 10) and decreased progressively thereafter.

Child questioned his results at the upper grade levels. He reasoned that flaws in procedures caused the irregular fluctuation. Older children had the opportunity to communicate freely among themselves prior to entering the experimental situation. They talked about the purpose of the experiment and about the appropriate response. Because of this communication among *Ss*, their choices were considered to be group rather than individual decisions.

Thus, the irregularity of choices in Child's study at the upper grade levels could have resulted from this situation rather than from a basic irregularity in preference for more difficult goals.

In the present experiment, careful control was exercised over communication among subjects. No subject in any grade had the opportunity prior to being tested of discussing the experiment with another member of his grade. Thus, choices were independently and individually determined. With this factor controlled, irregularity in preference for difficult goals was not found at the upper grade levels, neither when Child's two tasks nor when all four tasks were considered.

It seems probable that preference for a difficult goal is not a simple generalized trait at any age level. Choice of any individual may be influenced by a multiplicity of variables: How highly is the goal object valued? How strong is the barrier to the achievement of the goal and what form does the barrier take? What situational forces are acting upon the child? What is the sequence of activity? How does the child express his preference? Different patterns of interaction between these variables may exist at different age levels.

Summary

The main question under investigation was whether preference for the goal more difficult to attain is a generalized trait which increases in proportion to increasing age. Secondary questions were whether the sex of the experimenter and the type of barrier had a significant effect on preference.

Groups of 20 boys from grades 2, 4, 6, and 8 were closely equated on intelligence and socioeconomic status. Boys were assigned randomly to experimenters at each grade level. Each subject was confronted with four tasks in which he was asked to make a choice between a goal difficult or easy to attain. The barriers to the attainment of goals consisted of distance between the subject and the goal and of the number of pieces used to complete Formboard and Assembly tasks.

The first hypothesis was not supported. The percentage of boys choosing difficult goals under these experimental conditions increased from 32 per cent at grade 2 to 55 per cent at grade 4 but decreased to 39 per cent and 31 per cent at grades 6 and 8. No significant differences were found in preference for difficult goals between the individuals assigned to the male or female experimenter. The nature of the task and the type and strength of barrier used had a significant effect on choices of difficult goals.

This experiment suggests that preference for difficult goals is not a simple generalized trait which increases with age. At any age level preferences may vary in terms of a number of associated variables. Important factors are the intrinsic value of the goal object, the strength of the barrier, the motivating conditions within the experimental situation, the sequence of activity, and the manner in which choice is expressed by the child. Subsequent research

should consider the interaction of these variables and their influence on preference for easy- or difficult-to-attain goals.

References

1. Child, I. L. Children's preferences for goals easy or difficult to obtain. *Psychol. Monogr.*, 1946, 60, No. 4 (Whole No. 280).
2. Child, I. L., & Adelsheim, E. The motivational value of barriers for young children. *J. gen. Psychol.*, 1944, 65, 97-111.
3. Wright, H. F. The effect of barriers upon strength of motivation. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.), *Child behavior and development*. McGraw-Hill, 1943. Pp. 379-396. (Originally published in a monograph in 1937.)
4. University of Chicago. Methodological and operational considerations in the application of the Index of Status Characteristics to the metropolitan community, 1956.

EFFECT OF EXPOSURE TO SYMBOLIC AGGRESSION ON AGGRESSIVE BEHAVIOR [1]

O. Ivar Lövaas* [2]
University of Washington

This study was concerned with the effect of exposure to symbolic aggression on the play behavior of children. A bar-pressing response which produced aggressive doll action was observed immediately after the children had been exposed to either an aggressive or a nonaggressive motion picture.

Previous research on changes in aggressive behavior has been oriented primarily around testing the hypothesis of a cathartic or drive-reducing effect of aggressive expression (2). The data tend to support the hypothesis of drive-reduction when the aggressive expressions have been varied in the immediate experimental situation. Feshbach (5) observed a decrease in aggressive behavior immediately following the writing of aggressive TAT stories. Similarly Lewin *et al.* (10) observed a reduction in the over-all level of aggression in children subsequent to scapegoating behavior. When the experimental condition consisted of more passive exposure to aggressive behavior, such as viewing aggressive films, the reduction in aggressive behavior has not been demonstrated. Siegel (11) failed to find a difference in amount of aggressive play behavior in children who had been exposed to an aggressive film. This difference in findings is analogous to results concerning behavior for reinforcers based on the primary appetitive drives. If the organism does not consume foodstuff, but merely tastes it or sees it, his behavior to obtain foodstuff is increased rather than decreased (1, 7).

The assumption is made that the aggressive stimuli (of doll action) act as reinforcing stimuli. It is supported by Burns' data[3] which show that, when an aggressive action (in dolls) was contingent upon a response (leverpress), that response was strengthened. The remaining conceptual frame-

---

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] This study was supported by a grant from the National Institute of Health (M-2208).

[2] The author expresses his gratitude to Professors Donald M. Baer and Sidney W. Bijou for their continued assistance in this research. He also is indebted to staffs at the Seattle Day Care Nursery and at the Nursery School in the Gatzert Institute of Child Development, University of Washington, for their cooperation.

[3] Personal communication from R. C. Burns.

Child Develpm., 1961, 32, 37-44.

work for the study was oriented around research in relation to primary appetitive reinforcers. An attempt was made to investigate whether there is a change in aggressive behavior following exposure to aggressive stimuli analogous to the change in eating behavior following exposure to foodstuff. A similar conceptual framework was proposed by Gewirtz and Baer (6) in their study on the reinforcing properties of social approval. Their study demonstrated that the effectiveness of social approval as a reinforcer was related to satiation and deprivation operations in the manner analogous to such operations on primary appetitive reinforcers.

Experiment I

Method

The research was conducted in the Gatzert Institute of Child Development. Two experimental rooms were employed. The "playroom" contained the play apparatus, a screen for the presentation of the films, and a partition in one corner behind which *A* (an adult assistant) could sit. The other room, an observation room with one-way mirrors and sound equipment, contained the apparatus for presentation of the films and recording of behavior. The *E* (experimenter) could present the various experimental treatments from this room without being seen by the *S* (subject).

The films were presented by a Busch "Cinesalesman Continuous Projector" and appeared on a 9 by 12 inch screen in one wall of the playroom. A shutter blocked the lens every 10 seconds, for a period of five seconds. The *S*s were told that they could turn the film on when it went off by depressing a four-inch lever protruding from a small box on a table in front of the child. If the lever was pressed any time during the five-second off period, the shutter would reopen and allow for another 10 seconds of viewing before the next interruption. Having the *S*s "operate" the films in this manner assured continued attention to the films.

Two films were employed. The aggressive film (Official Films: "Rassling Match") was cut to give an almost continual display of aggression (hitting, biting, etc.) inflicted by one humanlike cartoon figure upon another. The nonaggressive film (Castle Films: "Bear Facts") depicted three bear cubs and a mommy bear engaging in pleasant humanlike play. The cut versions of each of these films had a running time of five minutes.

The doll apparatus was similar to that employed by Burns.[3] The dolls were six inches high, similar looking, dressed in boys' clothes, and facing each other about three inches apart. The doll on the right held a stick in its hand. They were placed inside a plexiglass box on top of a wooden box with a four-inch lever protruding from the front. Depression of this lever would activate a concealed motor which would swing the doll's arm with the stick rapidly down and back up into its original upright positition, making the stick hit the other doll on the head. Operation of the lever also activated an electric counter.

The subjects were children from the Institute's Nursery School. They were above average in intelligence and came from families with fathers in occupations above the means of the general population (3). Twelve children from the five-year-old group were selected in a random manner except for elimination of those whose cooperation was known to be difficult to obtain. They had considerable experience as *S*s in other experiments, but had no previous contact with the films or dolls employed in this study. Except for matching on sex, they were randomly divided into two groups of six each. One group was exposed to the aggressive film (AF group), the other group to the nonaggressive film (NAF group); otherwise, they were given the same experimental treatment. The *S*s were seen in the morning while they were engaged in free play. Both groups were run on any one day and matched on hour of the day. They were seen for three trials, spaced about four days apart.

A, a female adult selected because she was "good" with children, had visited the nursery prior to the experimental run and acquainted herself with the children. Each *S* was accompanied to the playroom by *A* who would invite the *S* to "play some games and look at a movie." (The playroom was located in the same building as the nursery.) When the *S* entered the playroom, he was made familiar with the room first and then introduced to the dolls. *A* pressed the lever once and said, "When you press this bar this (pointing) will happen. Now you sit in your chair and play with this while I sit in my chair." He received a two-minute period of continual reinforcement, i.e., one doll action for each bar press. The period was timed from the *S*'s initial depression of the lever. This period constituted the pre-experimental or operant level of doll action. *A* then placed the dolls out of the *S*'s view and said, "You will now see a movie. It will be shown right here (pointing to the screen). You are going to be the operator. That means that you have to turn the movie on when it goes off. Each time the movie goes off, press this lever (pointing to the box on the table) and the movie will come back on." The film was started and *A* made sure that the *S* knew how to turn the film on. The film continued for five minutes after which *A* told the *S* to play some more with the dolls. A two-minute period of continuous reinforcement followed, timed from the initial bar-press after the film. At the end of the session *S* received a "thank you" and was accompanied back to the nursery.

A interacted with *S* in a matter-of-fact manner. During the session she remained behind a partition except when she introduced the film or the dolls. She did not respond to *S*'s attempt to converse, unless he came over to her chair at which time she stated: "I'll sit here and you be over there for a few minutes more."

The second and third trials were identical with the first, with two exceptions. First, the operant period was eliminated. Second, doll play in the third period was immediately followed by a one-minute extinction period in which the second, fourth, and eighth response were reinforced to assure prolonged extinction. One *S* from each of the groups refused to return for

the third trial. In an attempt to prevent further loss of *Ss*, the reinforced doll play session in the third trial was shortened to one minute.

Results and Discussion

The differences in the mean number of responses on the dolls after exposure to the films were not statistically significant. The AF group showed fewer responses than the NAF group in the first postfilm session. This difference reversed itself by the last session (the extinction session). A second experiment with a larger *N* was undertaken in an attempt to obtain a statistically significant difference between the groups in the first postfilm session.

Experiment II

Method

This experiment was conducted in a trailer laboratory (4). The trailer was parked in the play yard of a day care center supported by the Community Chest. This school accepted children only from low income families with working mothers. The *Ss* were 20 children from four to six years of age. Except for matching on sex and age, they were randomly divided into two groups of 10 *Ss* each, forming AF and NAF groups analogous to those in Experiment I. None of the children had previous familiarity with the films or the dolls. The experimental procedure was identical to that in Experiment I, except that a four-minute extinction period followed immediately after the two-minute postfilm session of continual reinforcement. The experimental treatment was: a two-minute continual reinforcement, operant session; five minutes of the respective film; a two-minute period of continual reinforcement; a four-minute period of extinction with the second, fourth, and eighth response reinforced.

Results and Discussion

Despite the increase in *N*, none of the differences between the groups was statistically significant.

At the completion of Experiment II the experimental conditions were re-examined for possible sources of interference with the behavior following seeing the film. There was a possibility that the behavior involved in complying with instructions was stronger than the behavior resulting from the experimental conditions. In the preceding studies it appeared that some of the *Ss* persisted in pressing the lever for the doll action because they were told to do so by *A*. Many *Ss* showed signs of uneasiness about leaving the dolls during the latter part of the period, but they would continue pressing the lever without even attending to the dolls. The verbal instructions apparently kept the *Ss* operating the dolls, hence competing with and obscuring the effect of the films in controlling this behavior. This notion was investigated in Experiment III.

Experiment III

Method

The experimental setup was modified to minimize the interfering effect of "following instruction." Another toy was placed in the situation giving the *S* a choice for responding. This toy consisted of a lever and a wooden ball enclosed in a cage. Depression of the lever made the ball go to the top of the cage and back, through obstacles, to its original position. Depression of the lever also activated an electric counter. The boxes containing the dolls and the ball toy were placed side by side on a table, enabling *S* to work both simultaneously if he so desired.

The same *S*s used in Experiment II returned a week later but with reversed conditions, i.e., the group who had seen the aggressive film in Experiment II saw the nonaggressive film in Experiment III and vice versa. Any difference between the two groups in the response rate on the dolls would have to be attributed to the films, as operant level of responding was identical for the two groups.

When *S* came back to the trailer this time, he was introduced to the ball toy, shown how it worked, and asked to play with it for one minute. He was then reintroduced to the dolls and asked to press the bar to see if he remembered how it worked. He was allowed to press it five times and was then presented with the (different) film. After film viewing was completed, he was told, "You may play with this (pointing to the ball toy) or this (pointing to the dolls and lever), play with the one you like but start playing with this (pointing to the dolls and lever)." *S* was requested to play with the dolls first to insure that some behavior was recorded on the dolls. *A* waited beside *S* until he made his first response on the dolls, then retired behind the partition. It was intended that he should be free to leave the dolls whenever he wanted. If after two minutes he had not played with the ball toy, he was told, "You may play with either one, play with the one you want to play with." This session lasted for four minutes.

Results and Discussion

The total number of responses on the bar operating the dolls and the ball toy for the two groups are presented in Table 1 and an analysis of variance of these data is presented in Table 2.

The statistical tests show a significant interaction between the kind of toy played with (ball and doll) and the kind of film preceding the toys. When the *S*s could choose between the toys, the *S*s who had been exposed to the aggressive film engaged in more play behavior with the hitting dolls than did *S*s who had been exposed to the nonaggressive film. The AF group had almost twice as many responses on the toy with the aggressive doll action as the NAF group. This difference is statistically significant at the $p < .06$ level, one-tailed test ($t = 1.62$, $df = 18$). The variability between *S*s was considerable. Also the difference between the two groups was attenu-

TABLE 1

RESPONSES ON DOLLS AND BALL FOLLOWING AGGRESSIVE AND NONAGGRESSIVE FILMS

	Mean Responses on Toy		
	Dolls	*Ball*	*Sum*
Aggressive Film	98.2	99.5	197.7
Nonaggressive Film	58.6	142.1	200.7
Sum	156.8	241.6	

TABLE 2

ANALYSIS OF VARIANCE FOR RESPONSES ON DOLLS AND BALL FOLLOWING AGGRESSIVE AND NONAGGRESSIVE FILMS

Source of Variation	*Sum of Squares*	*df*	*Mean Sqare*	*F*
Between Films	22.5	1	22.5	
Between Toys	21529.6	1	21529.6	5.77*
Interaction: Films × Toys	16892.1	1	16892.1	4.53*
Within	134361.8	36	3732.2	
Total	172806.0	39		

* Significant at $<.05$ point.

ated by "obedient" *Ss*, i.e., *Ss* who were told at the end of two minutes of continuous doll play to play with either toy, immediately switched to the ball toy and stayed with that toy for the remaining two minutes without switching back to the dolls. There were two such *Ss* in the AF group, four in the NAF group. The four "obedient" *Ss* in the NAF group had a higher response rate than the remaining *Ss* in the NAF group. They obscured the difference between the groups by accentuating the over-all response rate of the NAF group.

The results of this study give evidence for an increase in responding to aggressive reinforcing stimuli as a function of exposure to symbolic aggression. From a theoretical viewpoint, exposure to symbolic aggression may have at least four effects upon subsequent aggressive behavior. First, the exposure may satiate the organism on aggressive reinforcers. Secondly, it may constitute an operation which increases the organism's generalized drive level (13). Thirdly, it may sensitize the organism, in the manner of Spence's K (13), by making behavior with aggressive consequences more probable. Fourth and lastly, the operation may provide discriminative stimuli, in the manner of Skinner's analysis (12), marking the occasion whether aggressive behavior can occur or not.

The first of these alternatives was not supported since there was an increase in behavior with aggressive consequences rather than a decrease which should occur under satiation. The second consequence, pertaining to increased generalized drive, similarly was not supported. Inspection of Table 1 indicates that the total means of the two groups on the two toys was almost identical (198 for the AF group, 201 for the NAF group). The higher response rate of the AF group on the dolls did not come about because of a generalized heightened activity level as a function of increased generalized drive, but rather because the aggressive film selectively increased behavior that had aggressive consequences.

The specific increase in responding for aggressive consequences is consistent with both alternative three and four. The third alternative, concerning increased sensitization to behavior with aggressive consequences, can be presented in Spence's framework (13) in the following manner. Availability of or exposure to a reinforcing stimulus can produce either an increase or a decrease in behavior for that reinforcer. If the exposure includes only a visual response to the reinforcing stimulus, K (incentive motivation) will be increased. An increase in K brings about an increase in probability of responding for that reinforcer.

The fourth alternative, change in the discriminative stimuli as a function of exposure to aggression, might be described in Skinner's framework (12) as follows. The sight of aggression, like the sight of food, constitutes a discriminative stimulus. For these children this particular film constituted a discriminative stimulus marking the occasion when aggressive behavior would be reinforced positively or at least not punished. A different film could mark the occasion when a child could not aggress successfully; hence it could constitute a discriminative stimulus for not aggressing.

The films employed in this study were not unlike those presented for children's entertainment, e.g., on TV, and the children did not represent an unusual sample of preschool children. There is a growing body of empirical evidence to support generalizations from play behavior to social behavior (8, 9). Hence the educational implication from this study is that aggressive films are likely to make children more aggressive rather than less aggressive.

Summary

Three experiments were performed on the effect of exposure to symbolic aggression by means of motion pictures on the play behavior of children. A bar-pressing response which produced aggressive doll action was observed immediately after the children had been exposed to either an aggressive or a nonaggressive film. The study gave evidence for an increase in responding for the aggressive doll action after exposure to the aggressive film when such responding is properly observed. The data are analogous to findings concerning behavior for reinforcers based on the primary appetitive drives. The

effects of this visual exposure to symbolic aggression can be conceptualized either in terms of an increase in incentive motivation (K) or in terms of providing discriminative stimuli marking the occasion when aggressive behavior will be reinforced or at least not punished.

References

1. Anderson, E. E. The externalization of drive. IV. The effect of prefeeding on maze performance of hungry non-rewarded rats. *J. comp. Psychol.*, 1941, 31, 349-352.
2. Berkowitz, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, 55, 257-284.
3. Bijou, S. W. Patterns of reinforcement and resistance to extinction in young children. *Child Develpm.*, 1957, 28, 47-54.
4. Bijou, S. W. A child study laboratory on wheels. *Child Develpm.*, 1958, 29, 425-427.
5. Feshbach, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 3-11.
6. Gewirtz, J. L., & Baer, D. M. Deprivation and satiation of social reinforcers as drive conditions. *J. abnorm. soc. Psychol.*, 1958, 57, 165-172.
7. Grindley, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, 20, 173-180.
8. Hollenburg, E., & Sperry, M. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, 1, 32-43.
9. Lebo, D., & Lebo, E. Aggression and age in relation to verbal expression in nondirective play therapy. *Psychol. Monogr.*, 1957, 71 (20).
10. Lewin, K., Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally-created "social climates." *J. soc. Psychol.*, 1939, 10, 271-299.
11. Siegel, A. E. Film-mediated fantasy aggression and strength of aggressive drive. *Child Develpm.*, 1956, 27, 351-365.
12. Skinner, B. F. *The behavior of organisms.* Appleton-Century-Crofts, 1938.
13. Spence, K. W. *Behavior theory and conditioning.* Yale Univer. Press, 1956.

SEX DIFFERENCES IN AGGRESSION AND ITS CORRELATES IN MIDDLE-CLASS ADOLESCENTS[1]

LEONARD M. LANSKY,* VAUGHN J. CRANDALL, JEROME KAGAN, *and* CHARLES T. BAKER[2]

Fels Research Institute for the Study of Human Development

Students of human development agree that adolescence is one of the crucial developmental periods. It is a time when the problems of dependence and autonomy with respect to parents and other adults are salient, and a period when sexual attitudes and social role identities are clarified. However, there is a very limited body of scientific information about adolescents, and even less rigorous theory. This paper presents some results concerning relations among a number of behavioral areas salient for normal, middle-class adolescents. The areas under study included aggression, achievement, dependence, independence, affiliation, identification, and moral standards. This report focuses on sex differences in the relation between aggressive behaviors and other behavioral areas.

METHOD

Subjects

The *S*s were 54 children between 13 and 18 years of age currently enrolled in the longitudinal study of the Fels Research Institute for the Study of Human Development. The sample was comprised of 32 boys and 22 girls. At the time of the first assessment contact, the boys ranged in age from 13-7 to 17-10; the girls, from 13-4 to 17-11. At the time of the second assessment session, the age-range of the boys was 13-10 to 18-7; of the girls, from

* Graduate School of Education, Harvard University, Cambridge 38, Massachusetts.

[1] This research was supported in part by research grant M-1260 from the National Institute of Mental Health, U. S. Public Health Service.

[2] Now at the University of Rochester School of Medicine.

Child Develpm., 1961, 32, 45-58.

13-5 to 18-8. The *Ss* were white and came from predominantly middle-class homes. Their parents were better educated than national averages. The mean Stanford-Binet IQ of the children was 113 with a standard deviation of 11.4.[3]

Assessment Procedures

Two contacts were made with each child. The first session included a semistructured individual interview lasting between one and one-half and two hours followed by a modified Rorschach test. For the second contact, the *Ss* were seen in small groups, and questionnaires and tests were administered.

Interview.[4] The interview schedule was designed to explore the following areas of behavior: achievement, dependence, affiliation, self-esteem, aggression towards parents and authorities, attitudinal differences with parents, anxiety over heterosexual relationships, and identification with parents and parental models. Three of the writers (LL, VC, JK) each interviewed approximately one third of the *Ss*. Each interview was electrically recorded with the *S*'s knowledge.

Modified Rorschach. Following the interview, each *S* was administered a 32-item modified Rorschach test. The *S*'s first response and response latency were recorded for each of the 32 stimuli. The modified Rorschach stimuli were constructed by placing a template of white cardboard over part of the standard Rorschach plate leaving a common *D* or *d* area exposed. A description of the stimulus areas appears in a separate report (8).

Group administered tests. The remaining tests were administered to the *Ss* in small groups (two to six subjects in a group) during the second assessment contact. Instructions to the *Ss* stressed: (a) the tasks were not tests in the sense that there were any right or wrong answers; and (b) no premium would be put on speed. Since all *Ss* began each test together, reading material was available for those who finished early.

In the order of administration, the tests were the Gough Brief Femininity Scale (7), the Franck Drawing Completion Test (5), the French Insight Test (6), a story completion test, and a self-rating inventory.

The Variables

The variables and instruments from which they were obtained are listed in Table 1. The variables are grouped by the behavioral areas of aggression, achievement, affiliation, dependency, sex anxiety, sex role identification, severity of standards, and defenses against guilt.

The interview ratings. Immediately following each interview, the interviewer rated the *S* with respect to 17 variables. Each scale used for these

[3] The IQ tests were administered by Dr. Virginia L. Nelson.

[4] A copy of the interview schedule and other assessment techniques mentioned in this report may be obtained by writing to the Department of Psychology, Fels Research Institute, Yellow Springs, Ohio.

TABLE 1

FINAL VARIABLES AND THEIR SOURCES

Variable	*Source*
Aggression	
Expressed criticism of father—I	Interview
Expressed criticism of mother—I	Interview
Reported aggression towards father—I	Interview
Reported aggression towards mother—I	Interview
Aggression—SRI	Self-rating Inventory
Aggression—R	Rorschach
Achievement	
Anxiety about lack of achievement—I	Interview
n Achievement—Fr	French Test
n Achievement-recognition—Fr	French Test
Dependence	
Conformity to authority—I	Interview
n Dependence—Fr	French Test
n Autonomy—Fr	French Test
Affiliation	
n Affiliation—Fr	French Test
Acceptance—SRI	Self-rating Inventory
Sex Anxiety	
Latency to male cards—R	Rorschach
Latency to female cards—R	Rorschach
Sex Role Identification	
Identification with father—SRI	Self-rating Inventory
Identification with mother—SRI	Self-rating Inventory
Franck (masculinity-femininity)	Franck Test
Gough (masculinity-femininity)	Gough Femininity Scale
Severity of Standards	
Severity of standards—SCB	Story Completion—Boxes
Severity of standards—SCP	Story Completion—Photography
Defenses against Guilt	
Denial SCB (SCP)*	Story Completion Item
Intellectualization SCB (SCP)*	Story Completion Item
Accept responsibility SCB (SCP)*	Story Completion Item

* Each variable was considered independently for the boxes story (SCB) and the photography story (SCP).

ratings contained an introductory description of the variable and a six-point scale with cue-point referents for the extreme points (1 and 6) and for scale points 3 and 4.

Thirty-three interviews were randomly selected for the determination of interrater reliability, 11 from each of the three interviewers. The measure of interrater reliability was the percentage of essential agreement between each interviewer and an independent rater (CB). The latter made his ratings after listening to electrical recordings of the interviews. Two ratings were said to agree if they were within one point of each other. The criterion for acceptable reliability for each variable was a percentage agreement of 80 or better.

Six of the 17 variables rated met the dual criteria of (a) satisfactory interrater reliability and (b) sufficient range to be considered worthy of further analysis.[5] The six interview variables were: (a) *expressed criticism of father,* (b) *expressed criticism of mother,* (c) *reported aggression toward father,* (d) *reported aggression toward mother,* (e) *anxiety associated with lack of achievement,* and (f) *conformity with authority.*

Modified Rorschach. The variables obtained from this instrument included: (a) *latency to phallic stimuli*—for each *S,* the latencies to the 32 stimuli were rank ordered, and the average rank for the four stimuli which were strongly suggestive of a phallus comprised this score; (b) *latency to vaginal stimuli*—four of the ink blot stimuli were suggestive of female genitals, and the latency score was the average rank for these four cards; (c) *aggression*—this score consisted of the number of responses in which people or animals were described in aggressive activity (people fighting, animals struggling), a destructive scene was perceived (landslide, A-bomb), or people, animals or objects were seen as dead or mutilated. Interrater reliability for this score was 92 per cent.

Gough Brief Femininity Scale. This scale is a 58-item self-rating test. For such items as "I want to be an important person in the community," the subject indicates whether the statement is true or false. The final score consists of the number of items where the response is coded feminine. The scoring code was derived from previous empirical studies (7). The test covers several aspects of masculinity-femininity including preferences for conventional occupational roles, sensitivity to nuances of social relationships, and concern with one's own feelings and one's own body.

Franck Drawing Completion Test.[6] This measure, in contrast with the Gough test, was designed to tap more covert levels of masculine vs. feminine

[5] Interview variables which did not show sufficiently broad distributions were: manifest hostility during the interview, differences with father's attitudes and values, differences with mother's attitudes and values, and opposite sex interests. Variables with unacceptable reliability were: manifest dependency in the interview, affiliative needs with same sex peers, anxiety over heterosexual activity, importance of achievement, and tendency to behave dependently with others.

[6] The copyright of this test is owned by Mr. S. S. Dunn, Bureau of Educational Research, Melbourne, Australia.

identification. The Franck Test consists of 36 line drawings to be completed by the *S* in any way he wishes. The score consists of the number of items coded feminine on the basis of either the content of the drawing or its formal properties. For example, expanded drawings are scored masculine, while internally elaborated ones are scored feminine. According to Franck and Rosen (5), the test assesses unconscious levels of masculinity-femininity, levels which are hypothesized to be related to body image of the *S*. In a previous investigation (9), the coder of the present study (LL) demonstrated 90 per cent essential agreement using the Franck scoring manual (4).

French Insight Test. This test was used to assess the *S*'s preoccupation with several needs or goals. As employed in the present study, the test consisted of the 20 items used in the two forms of the original test (6). There were separate forms for boys and girls. Each item described a characteristic behavior of a boy (girl), e.g., "Joe would rather follow than lead," "John said, 'Look what I've done'," "Tom never joins clubs or social groups," and the *S* must describe the personality of the protagonist. For each item the *S* answered three questions: (a) what is the boy (girl) like? (b) what does he (she) want to have or do? (c) what are the results of his (her) behavior apt to be? The *Ss*' responses to the test items were scored for preoccupation with the following needs or goals: *achievement,* when the *S*'s description of the stimulus character's behavior indicates a goal of approval (from self or others) for intellectual or vocational competence; *affiliation,* when the goal is to be liked or accepted by others or to be part of a group; *dependence,* when the goal is to obtain help or emotional support from others; *autonomy,* when the goal is to be self-sufficient, to avoid direction, manipulation, dominance, or help from others.

For each of the 20 test items, if none of the above-mentioned needs was predominant, a score of zero was given. Otherwise, the predominant need was scored. When achievement preoccupation was scored, a subscore was given for achievement-recognition if the *S* clearly stated that the major motivation for achievement was recognition from other persons.

One story at a time was rated for all *Ss* before the next story was assessed, and the *Ss*' identities were disguised. Interrater reliability was determined on an independent sample and exceeded 80 per cent essential agreement for each of the variables rated.

Story Completion Test. This test, a projective technique, was designed to elicit responses regarding severity of moral standards and defenses against guilt following transgression of such standards. There were separate forms for boys and girls. The test consists of three incomplete stories, each of which is about a young boy (girl) who violated an established and accepted moral standard in relationship to: (a) his parents (story 1), (b) his peers (story 2), and (c) his male employer (story 3). The *Ss* were instructed to complete each story by telling how the hero feels, what happens, and how the experience turns out. This story technique and the types of code used have been described in recent publications elsewhere (1). Only the first and third

stories were coded in the present study. The first incomplete story, hereafter referred to as "the boxes story" or SCB, concerned disobedience toward parents. Despite previous orders to the contrary, the central character takes some boxes down from his parents' closet. The third incomplete story, hereafter called "the photography story" or SCP, described the central character's destroying valuable photography equipment belonging to his male employer after the latter frustrates him. In both stories, the theme clearly implied that no one knew, or could find out about, the hero's (heroine's) actions.

Each story was coded for the presence or absence of several defenses against guilt. These included: (a) *denial*—denial was scored when the *S*'s response contained events which contradicted those of the story beginning; (b) *intellectualization*—this variable was scored for responses which were essentially devoid of any expression of affect; (c) *acceptance of responsibility*—this variable was scored when the hero (heroine) acknowledged responsibility for his (her) aggression and violation of a moral standard and paid for the violation in a manner that "fitted the crime."

Each story was also rated for *severity* of moral standards on a seven-point scale. This rating evaluated the degree to which the hero (heroine) punished himself, consciously or unconsciously, for his actions. One story for all *S*s was coded at a time, and all scoring was done with the *S*'s identity disguised. The coding was checked on an independent sample and found to exceed 85 per cent essential agreement for each of the four variables coded.[7]

Self-rating Inventory. This 110-item instrument was designed to assess a number of behavioral areas. Relevant areas for the present study included: (a) *aggression,* (b) *desire for acceptance,* (c) *identification with mother,* and (d) *identification with father.* The *S* was asked to place a number from 1 to 4 by each statement indicating the degree to which he felt the statement was characteristic of himself. The inventory contained 16 aggression items which included statements such as: "When I get mad at somebody, I usually tell them so"; "I get mad when someone tells me to hurry with some job." The Inventory's six acceptance items were: "I don't think you have to have friends in order to be happy"; "I'd like to be able to make friends easier"; "Occasionally I'll go along with the crowd, even though I don't want to, because I don't want to be different"; "I find it hard to say 'no' when people ask me to do something"; "I try hard to make new friends"; "I sometimes feel lonely and left out of things." The five identification with mother items were: "I agree with most of the ideas and attitudes of my mother"; "When I was a child my mother was pretty strict with me"; "I often wish my mother were smarter than she is"; "I sometimes think that I'd like to be more like my mother in temperament and personality"; "I feel that I know a little more about life than my mother." The five identification with father items were: "I agree with most of the ideas and beliefs of my father"; "I often wish my father were more successful than he

[7] The authors are grateful to Paula Brown, a senior in the Department of Psychology at Antioch College, for doing the reliability check.

is"; "I feel that I know a little more about life than my father"; "When I was a child, my father was pretty strict with me"; "I sometimes think that I'd like to be more like my father in temperament and personality." The *Ss*' score for each variable was the average of his ratings for the items pertaining to that variable.

Statistical Analysis of the Data

For each variable from each assessment technique, the distribution of scores was divided into high and low at the cutting point closest to the median score for the total sample of *Ss*. This division was used for all statistical tests of sex differences. Median splits were obtained for the boys and girls separately for analyses of relations between variables. Exceptions to the above were the scores derived from the French Test and the Story Completion Test: for these tests, the terms "high" and "low" refer to the presence and absence of the variable under consideration. When more than 40 *Ss* were involved in an analysis of the data, chi square corrected for continuity (11) was used; where there were fewer than 40 *Ss*, the Fisher-Yates exact test was employed (3, 10). All probabilities presented in the following data summaries are for two-tailed tests.

Results and Discussion

Sex Differences on Individual Aggression Variables

There were six variables concerned with aggression in the study. Four were interview variables: expressed criticism of father, expressed criticism of mother, reported aggression towards father, and reported aggression toward mother. The remaining two were aggression in the Self-rating Inventory and aggressive ink blot percepts.

Table 2 presents all the sex differences which were significant beyond the .10 level for a two-tailed test. In every case the statistic used was a cor-

Table 2

SEX DIFFERENCES ON INDIVIDUAL VARIABLES

Variable	*Direction*	*p**
Reported aggression toward father—I	B > G	< .05
Aggression—SRI	B > G	< .10
n Independence-autonomy—Fr	B > G	< .10
n Affiliation—Fr	G > B	< .005
Anxiety about sexuality—SRI	G > B	< .10
Gough Femininity Scale	G > B	< .005
Accept responsibility—SCB	G > B	< .005
Severity of standards—SCP	B > G	< .05

* The statistic used was chi square corrected for continuity. The probabilities are for a two-tailed test.

rected chi square. Only two of the eight involve aggression measures. Boys were higher than girls on reported aggression to father in the interview and on self-rated aggression.

Boys also scored higher than girls on independence-autonomy responses to the French Test and on severity of guilt on the Story Completion Test. In contrast, more girls than boys obtained high scores for need for affiliation as scored from the French Test and for self-ratings of anxiety about sexual activity. The girls also obtained higher scores on feminine attitudes and interest than boys and on acceptance of responsibility for destructive action in the Story Completion Test.

Most of these findings are congruent with the generally-accepted picture of sex differences in middle-class American adolescents. Adolescent boys are expected to be more aggressive than girls and to admit that they are. Boys are expected to be more preoccupied with the development of autonomy and independence while girls are more often pictured to be more concerned with affiliation and anxious about sexual activity. However, boys and girls were not significantly different from one another on four of the variables concerned with aggression: amount of expressed criticism of mother, amount of expressed criticism of father, reported aggression toward mother, and aggressive imagery on the Rorschach. As will be noted below, there were various sex differences both in the interrelationships among these aggression variables and the relationships between these aggression variables and variables assessing other behavioral areas.

Sex Differences in Relations between Aggression Variables

The relations between pairs of the six aggression variables for each sex are presented in Table 3. For these six variables, nine of the 30 exact tests reached the .10 level of significance or better for a two-tailed test. Six of these involved the boys; three, the girls.

TABLE 3

SIGNIFICANT RELATIONS AMONG AGGRESSION VARIABLES

	Expressed Criticism of Father—I		Expressed Criticism of Mother—I		Reported Aggression to Father—I	
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Expressed Criticism of Mother—I . .	+.004					
Reported Aggression to Father—I . .	+.02	+.05	+.02			
Reported Aggression to Mother—I .			+.01	+.02	+.02	+.03
Aggression—Rorschach			+.08			

NOTE.—The cell entries are two-tailed p values for the exact test. The plus signs indicate that all significant relations were positive.

The three significant relations for the girls involved the same pairs of variables as three of those for the boys. For both boys and girls, reported aggression to father and mother in the interview were positively related to each other. There was also a significant relation, for both sexes, between interview ratings of expressed criticism of each parent during the interview and reported aggression towards that parent.

The boys were unique in that all of their additional significant relations were associated with expressed criticism of mother during the interviews. This variable was positively related to criticism of father in the interview, reported aggression toward father in the interview, and the amount of aggression acknowledged in the Self-rating Inventory.

The expressed criticism of, and reported aggression toward, each parent, which were positively related for both boys and girls, may be two manifestations of a more general propensity toward acceptance or nonacceptance of parents during the struggle for independence typical of American middle-class adolescence. The three additional significant relations for boys may be signs of the more intense struggle for independence by adolescent males, the greater overt aggression permitted boys in middle-class society, and/or more subtle sex differences in adolescent development.

Sex Differences in Relations between Measures of Aggression and Other Variables

Of the 264 exact tests run between aggression and other variables, 23 were significant at beyond the .10 level for a two-tailed test. However, 11 of the 23 were significant at the .02 level or beyond. The data are summarized in Table 4.

Contrary to the picture for relations among aggression variables, fewer significant relations of aggression variables to nonaggressive variables were found for boys than for girls. Sixteen of the 23 were for girls; seven, for boys. The patterns for boys and girls were also dissimilar. For the boys, the focal variable was aggressive ink blot percepts; for the girls, significant relations centered around expressed criticism of, and reported aggression toward mother.

For the boys, aggressive imagery on the Rorschach was *positively* related to high feminine scores on the Franck Test and to dependency on the French Test and *negatively* related to interviewer ratings of conformity to authority. Also, reported aggression toward father was associated with self-ratings of a strong need for acceptance and with preoccupation with dependency on the French Test. It is possible that the more aggressive boys of the present sample feel free to express aggressive and dependent wishes; or, it may be that their aggression is in reaction to passive-dependent "feminine" needs.

There were two other relations between aggression and other variables for the boys. Reported aggression toward mother in the interview was positively related to achievement preoccupation on the French Test; the self-

TABLE 4

SIGNIFICANT RELATIONS BETWEEN AGGRESSION AND OTHER VARIABLES

Other Variables	Expressed Criticism of Father—I		Expressed Criticism of Mother—I		Reported Aggression to Father—I		Reported Aggression to Mother—I		Aggression —SRI		Aggression —R	
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Achievement												
Anxiety about lack of Achievement—I	..	..	..	−.004	..	..	..	..	..	..	..	..
n Achievement—Fr	..	..	..	+.02	..	..	+.04	+.08	..	..	..	..
n Achievement-recognition—Fr	..	..	..	+.02	..	..	..	+.08	..	..	..	..
Affiliation												
n Affiliation—Fr	..	..	..	..	..	..	..	..	..	..	..	−.05
Acceptance—SRI	..	..	..	−.02	+.002	..	..	−.008	..	+.02	..	..
Dependence												
n Dependence—Fr	..	..	..	..	+.10	..	..	..	..	..	+.02	..
Conformity to authority—I	..	−.02	..	..	..	..	..	..	..	..	−.003	..
Identification												
Identification with mother—SRI	..	−.08	..	−.06	..	..	..	..	..	..	..	..
Franck Test	..	..	..	..	..	..	..	..	..	..	+.08	..
Sex												
Latency to female cards—R	..	..	..	−.05	..	..	..	..	..	..	..	..
Severity of Standards												
Severity—SCB	..	..	..	..	..	..	..	..	..	+.07	..	..
Severity—SCP	..	..	..	−.05	..	..	..	..	−.02	..	..	..
Defenses vs. Guilt												
Intellectualization—SCB	..	..	..	+.07	..	..	..	..	..	..	..	..

NOTE.—The cell entries are two-tailed test *p* values for the exact test. The plus and minus signs indicate the direction of relationship.

rating of aggression was negatively related to severity of standards on the photography story of the Story Completion Test.

For the girls, eight of the 17 significant results presented in Table 4 concerned relations of other variables with the girls' expressed criticism of mother during the interview. Girls who expressed relatively strong criticism toward their mothers were rated low in anxiety about achievement in the interview. They were slow to respond to Rorschach stimuli resembling female genitalia and rated themselves as being dissimilar to their mothers and having little interest in acceptance from others. In the Story Completion Test they expressed little guilt over aggressive behavior toward an authority figure (SCP); they also expressed little affect in the story about the violation of parental rules (SCB). Girls who were critical of their mothers were also preoccupied with achievement and recognition for achievement on the French Test. (For the girls only, achievement preoccupation and achievement recognition themes on the French Test were positively related at the .02 level.[8])

Three of the other significant relations for girls concerned reported aggression towards mother in the interview. Reported aggression toward mother was positively related to achievement preoccupation and to achievement recognition on the French Test and negatively related to need-acceptance on the Self-rating Inventory. These findings were parallel to those obtained for expressed criticism of mother in the interview.

There were five additional significant relations between aggression variables and other behaviors for the girls. Two of them involved expressed criticism of father in the interview. Girls who were critical of their fathers were rated low in conformity to authority by the interviewers and obtained low scores on identification with mother on the Self-rating Inventory.

Two of the three remaining significant results for the girls concerned the aggression score derived from the Self-rating Inventory. Girls who rated themselves as aggressive on the Inventory evidenced strong moral standards regarding disobedience to parental orders in the Story Completion Test, and girls who rated themselves high on aggression also rated themselves high on need for acceptance. Finally, a negative relation between the girls' Rorschach aggression score and preoccupation with affiliation on the French Test was obtained.

In summary, the results present an interesting pattern. For individual variables, there are fewer sex differences in aggression than predicted. For relations among aggression variables, there are more significant results for boys than girls, a finding which confirms the expectation that boys are more aggressive than girls during adolescence. However, the associations between the aggression variables and other behaviors give a different picture. There are more such significant associations for girls than for boys, and the patterns for the two sexes are strikingly different. Only two variables, reported

[8] This result does not appear in the tables of this paper.

aggression toward mother in the interview and achievement preoccupation on the French Test, are significantly related for both boys and girls. While for boys most of the correlations involved aggressive imagery on the Rorschach, the major emphasis for girls was on criticism of and aggression towards mother. It was suggested that the aggressive boys feel free to express various motivational content on a variety of instruments or that their aggression is in reaction to unacceptable dependent needs. For the girls, the aggressive responses towards mother were related to "masculine" interests, behaviors, and needs.

While the meanings of the results obtained are not clear, they are in accord with Douvan's suggestion (2), based on her study of a national sample of Boy and Girl Scouts, that sex differences in American adolescents may not be so much a matter of differences in individual variables per se as in the specific types of conflicts and methods of resolution unique to each sex. In the present study the patterns of personality characteristics associated with any one aggressive variable are different for boys and girls. These differences most likely reflect the many meanings of the variables we chose to call aggressive and the fact that aggressive behavior has different meanings, outlets, motives, and defenses associated with it for boys and girls.

Summary

This report presented data from a study of normal middle-class adolescents. The results reported focused on sex differences in relations between aggressive behavior on the one hand and achievement, dependence, independence, affiliation, sex anxiety, identification, and moral standards on the other.

The subjects were 32 boys and 22 girls enrolled with their families in a longitudinal study at the Fels Research Institute for the Study of Human Development. These data were obtained from individual interviews and from scores derived from several individual and group-administered assessment techniques. These techniques included a modified Rorschach, the Gough Brief Femininity Scale, the Franck Drawing Completion Test, the French Test, a specially constructed story completion test, and a self-rating inventory.

Major results were as follows: (a) On individual variables, boys scored higher than girls with respect to reported aggression towards father in the interview, aggression on the Self-rating Inventory, independence-autonomy on the French Test, and severity of moral standards on a story completion item. The girls scored higher than the boys with respect to preoccupation with affiliation on the French Test, anxiety about sexual activity expressed in the Self-rating Inventory, femininity on the Gough Test of feminine attitudes and interests, and in taking responsibility for aggressive action toward a frustrating authority as assessed from the Story Completion Test.

(b) There were more significant interrelations among aggression variables for the adolescent boys than for the girls. Significant findings for girls were limited to relations between interview ratings of expressed criticism of, and reported aggression towards, parents. For the boys, the aggression variables of the interview related to one another as well as to aggressive Rorschach imagery. (c) There were more significant relations for girls than for boys between their aggression scores, on the one hand, and measures of the other behaviors, on the other. For the boys, both aggression towards father and Rorschach aggression were related to self-rated need for acceptance and to dependency reflected in their projective test responses. For the girls, expressed criticism of mother was related to preoccupation with recognition for achievement, anxiety about sexuality, low concern with affiliation, low desire to be similar to mother, and minimal guilt about aggressive acts directed against a male authority. Expressed criticism of father by the girls was negatively related to desire to be similar to mother and conformity to authorities. In addition, for these girls, aggressive Rorschach imagery was negatively related to preoccupation with affiliation in the French Test responses.

Implications of the results with respect to the different sex role demands and conflicts of middle-class adolescent boys and girls were discussed.

References

1. Allinsmith, W. The learning of moral standards. In D. R. Miller & G. E. Swanson, *Inner conflict and defense*. Holt, 1960. Pp. 141-176.
2. Douvan, E. Character processes in adolescence. Paper read at Amer. Psychol. Ass., New York, September, 1957.
3. Finney, D. J. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika*, 1948, 35, 145-156.
4. Franck, K. Franck Drawing Completion Test and Manual. Unpublished ms., Australian Council for Educational Research, Melbourne.
5. Franck, K., & Rosen, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, 13, 247-256.
6. French, E. G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: a method of assessment and study*. Van Nostrand, 1958.
7. Gough, H. G. Identifying psychological femininity. *Educ. psychol. Measmt*, 1952, 12, 427-439.
8. Kagan, J. The long term stability of selected Rorschach responses. *J. consult. Psychol.*, in press.

9. LANSKY, L. M. Sex identity and defenses against aggression. In D. R. Miller & G. E. Swanson, *Inner conflict and defense*. Holt, 1960. Pp. 272-288.

10. LATSCHA, R. Tests of significance in a 2 × 2 contingency table: extension of Finney's table. *Biometrika*, 1953, 40, 74-86.

11. SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, 1956.

EXPERIMENTAL DEVELOPMENT OF "LIKE" AND "DISLIKE" OF OTHERS AMONG ADOLESCENT GIRLS

EVAN R. KEISLAR*
University of California, Los Angeles

Several studies have indicated that adolescent girls who are well liked by their peers are rated as "friendly" (1, 2, 4). But such descriptive studies do not in themselves demonstrate that girls like one of their peers because she is friendly. Assuming that friendliness means the expression of social approval or "like," to establish this cause and effect relationship we must somehow contrive to have girls liked by others and then note the effect upon the girls' reciprocal liking for these others.

This problem of the exchange of social reinforcements between persons has been studied by Sidowski, Wyckoff, and Tabory (5). In a minimal social situation, two subjects in separate rooms alternately gave each other either a score or a shock, a positive or a negative reinforcement respectively, by pressing one of two buttons. Even though neither subject was told which button resulted in a score for the other, the subjects learned to give each other a high proportion of positive reinforcements.

However, the problem posed by the present investigation involves a situation where the prior learning of the subject is clearly more relevant than in the study just cited. Here the task is to find out whether adolescent girls have already learned to give positive reinforcements, i.e., expressions of "like," to those who have previously shown a liking for them. The general design involved showing subjects that they were liked by one girl and disliked by another. The effect upon their expression of "like" to each of these other two girls was then noted.

---

* Department of Education, University of California, Los Angeles 24.

Child Develpm., 1961, 32, 59-66.

One could, of course, view this expression of "like" as the result of a change in attitude toward others or a shift in a preference for others, a change which occurred when the subjects were told that they were liked or disliked. But, in this study at least, the existence of such an attitude or preference can be inferred only on the basis of the subject's behavior, i.e., his expression of "like" or "dislike." One cannot use these terms, an attitude or a preference, to "explain" this behavior. To avoid such circularity of reasoning, the expression of "like" or "dislike" is here considered simply as behavior which the subjects have previously learned to exhibit under such circumstances.

This prior learning of the subjects may be viewed more profitably within the general theoretical framework presented by Skinner (6) on verbal behavior. The expressions of "like" or "dislike" may be regarded as classes of verbal behavior of a speaker. If a speaker's verbal responses of "like" are reinforced by a listener with approval, an approving listener comes to control such expressions of "like." In other words, individuals will more often show their liking of persons who have previously approved of them.

On the other hand, if the listener does not reinforce the speaker's expression of "like" or if the listener shows disapproval, indifferent or disapproving audiences less often arouse in the speaker such verbal responses of "like." In fact, if the speaker has on some occasions shown some form of "dislike" toward a disapproving listener, e.g., "go away" or "I don't like you," and this has resulted in the removal of this source of negative reinforcement, the speaker will learn to show his dislike of persons who disapprove of him. (Other types of verbal behavior such as submissiveness or attempts to "smooth the feelings" of listeners may, of course, be learned where this kind of behavior has resulted in the reduction of the listener's disapproval.) If we assume that the processes described are the dominant ones in the subjects' past history, this line of reasoning suggests that the "approving-disapproving" characteristic of an audience partly determines the extent to which a speaker shows "like." It was the hypothesis of the present study that the subjects, viewed here as speakers, would express a greater amount of "like" toward a girl, a listener, who had previously liked them than toward a girl, another listener, who had previously disliked them.

In this investigation three adolescent girls were tested at a time in a laboratory situation where the identity of each girl was hidden from the others to eliminate the extraneous effects of appearance, prior acquaintanceship, or irrelevant behavior. The girls answered questions on their preferences and expressed "like" or "dislike" of each other by pressing buttons which, they were informed, turned on appropriate lights the others could see. Conversely, each girl saw lights which presumably revealed the responses, answers to preference questions and expressions of "like" and "dislike," of the others.

In actuality there was no communication whatsoever among the girls. The lights each girl saw were really operated by the experimenter ac-

cording to a predetermined schedule. Each girl's responses were recorded but never were revealed to the others. In this way it was possible for the experimenter to control the administration of "like" and "dislike" given to each girl and hence ascertain the effect upon her own expression of "like" for the others.

Since the experimenter controlled all the information given, the three girls tested at one time were regarded as three separate subjects and received identical treatments. During the first part of the experimental session, each subject responded to a set of preference items. She was informed by means of lights, which the experimenter operated, that one of the other "girls" liked her on the basis of the way she answered the items and that the other "girl" disliked her. Following this treatment, each subject then expressed either "like" or "dislike" on the same basis to first one and then the other of the two "girls."

Hypothesis

Adolescent girls express a greater liking for a girl who has previously shown a liking for them than for a girl who has previously shown a disliking for them, such liking being defined as the number of times an individual indicates "like" of the other instead of "dislike" on the basis of the way the other answered a preference test item.

Method

Subjects

There were 72 ninth-grade girls with Otis scores over 100 enrolled in a suburban junior high school. The subjects were tested three at a time making 24 sessions in all.

Apparatus

Three booths, each four feet square, were placed side by side with open fronts so that each subject could see items which were projected through the center booth on the screen in front. In each booth in front of the subject was a small table on which were placed four button-type switches. Two of these buttons were used to indicate answers to the two-choice items projected; the other two buttons were used to indicate "like" or "dislike." A panel of four lights on the table revealed to each subject what button she had just pressed. The numbers "1" and "2" were illuminated in this panel for item answers; a green light and a red light in the panel were turned on for "like" and "dislike," respectively. The button-pressing responses of each subject closed relays from which this information was recorded by an IBM key punch after each item.

At each side of each booth were four more lights, the numbers "1" and "2" and a green and a red light, corresponding to the panel on the table. Subjects were informed that the four lights on the left of the booth were

turned on by the girl on her left (or at the other end of the row); similarly, the four lights at the right presumably indicated the responses of the girl on her right. All lights at the side were controlled, through a bank of relays, by means of a prepunched card read by the key punch. In this way each of the three subjects always observed the same pattern of lights as did the others, a schedule predetermined by the experimenter.

Since each time a girl pressed a button a relay closed with an audible click in the experimental room, it was necessary to mask the fact that the three girls were all answering at the same time. To accomplish this, a special relay was made to close and open irregularly throughout the experimental session, thus providing a continuous random sequence of audible clicks.

Test Items

All subjects were shown on the screen three series of two-choice test items adapted from Edwards' Personal Preference Schedule (3). Each series consisted of 26 items an example of which is given below:

1. I like to make new friends.
2. I like to be successful in whatever I undertake.

When responding to these items, subjects were instructed to select the statement which more nearly applied to them and to press the corresponding button.

Two forms of this test were constructed, each form consisting of 26 items. For half the subjects form A was used for the first series of items and form B for both the second and third series. The other half of the subjects were shown form B for the first series and form A for the second and third. For all subjects, therefore, the sets of items presented during series 2 and 3 were identical.

Procedure

Three girls were called out separately from different classrooms and were seated, one in each booth, without their seeing each other. Subjects were instructed to remain quiet and to ask no questions except in writing or by sign language. Even after the experimental session was over, the girls were sent back separately to their classrooms without learning each other's identity. This was done to prevent any possible embarrassment and to keep the girls from comparing notes.

Following an orientation set of six items, during which the subjects learned the purported functions of the buttons and the lights, the three girls were instructed as follows:

Three series of preference items, for which there were no right or wrong answers, would be projected on the screen in front. Each of the three girls

would in turn give her answers first for one series of items. For each item, the answer given by the girl responding first would be shown by means of lights to the other two girls. The number "1" or "2" would be lighted on the side of the booth. The other two girls would then indicate, by pressing either the "like" or the "dislike" button, whether they liked or disliked the first girl "on the basis of the way she answered the item." Subjects were instructed that this judgment of like or dislike was not to be made simply because they agreed or disagreed with the girl's answer; it was pointed out that a person might like someone even though she would not answer in the same way, and, conversely, she might dislike someone who gave the same answer as herself. When the two girls had made their judgment of "like" or "dislike," the first girl would be informed of the decisions of the others by means of the red and green lights on each side of the booth. After a few seconds all lights would be turned out, a new item would appear, and the process would be repeated.

When the first series of items was completed, another girl would answer first and would be informed of the "likes" and "dislikes" of the others. For the last series of items the third girl would answer first. So that each girl could find out for which series of items she was to answer first, subjects were instructed to look at a card on their tables where this information was presented.

In actuality, since these three order cards were identical, all three girls followed exactly the same order. During the first series of items, each subject gave her responses first to the preference items. (A yellow light was flashed in her booth if she took more than 10 seconds to respond.) She was then shown, after a five-second interval, a red light or a green light, one on each side of the booth. These lights, controlled by a prepunched card, informed each subject that one girl liked her on 23 of the 26 items and that the other girl disliked her on 23 of the 26 items.

The three subjects on the remaining two series of items indicated their "like" or "dislike" of each of the "other two girls" in turn. During the second series of items each subject was shown lights on one side of the booth, purportedly the responses of the girl on that side (or at the other end of the row) to the preference test items. During the third series of items subjects were shown lights on the other side of the booth. The sequence of the illuminated numbers for series 2 was made identical with that in series 3 by using the same punched card for both series. Since the same 26 items were shown for the second and third series, as has been previously explained, and since the pattern of lights shown on one side of the booth was identical with that shown on the other, the presumed preferences of the "other two girls" were therefore identical. It was assumed consequently that any difference in the number of "likes" given by each subject to the "other two girls" would be attributable entirely to the difference in treatments received from "them" in the first series.

For half of the subjects, i.e., for 12 of the 24 groups of three subjects each, the "girl" who liked them was on the left (at least "her" answers and expressions of "like" and "dislike" appeared on the left) while the "girl" who disliked them was on the right; the opposite was true for the other half. For half of each of these sets of groups the subjects were shown the answers of the "girl" who liked them on series 2 and those of the "girl" who disliked them on series 3; for the other half of each of these sets of groups the subjects saw the answers of the "girl" who disliked them on series 2 and those of the "girl" who liked them on series 3. Finally, for each of these subsets, half of the groups of subjects were shown form B of the preference test for series 2 and 3; for the other half, form A was used for these last two series.

Results

The mean number of "likes" given to the girl who liked the subject previously was 19.1 with a standard deviation of 2.8. The comparable figure for the girl who disliked her was 17.7 with a standard deviation of 2.3. The correlation between the two measures, based on 72 cases, was .53. To test the significance of this difference an analysis of variance was carried out analyzing the total variance into "within" and "between" subjects. The factors were the two tests of number of "likes" obtained from the last two series (T), two positions, right or left of the booth for display of the responses of the girl who liked them (P), two orders of presenting the other

Table 1

ANALYSIS OF VARIANCE OF NUMBER OF "LIKES" GIVEN TO TWO GIRLS, ONE APPROVING, THE OTHER DISAPPROVING

(Given by 72 adolescent girls)

Source	*df*	*Mean Square*	*F*
Between *Ss*	71		
P (Position)	1	6.6	. . .
O (Order)	1	25.8	. . .
P × O	1	73.8	5.8*
Error (b)	68	12.8	. . .
Within *Ss*	72		
T (Tests)	1	70.8	16.1**
T × P	1	.1	. . .
T × O	1	.6	. . .
T × P × O	1	6.6	. . .
Error (w)	68	4.4	
Total	143		

* Significant at .05 level.
** Significant at .001 level.

girls' responses (O), with 18 subjects in each subgroup. The effect of the different forms was not analyzed. Since nine subjects in each subgroup were shown form A for the last two series and the other nine in each subgroup form B, this factor was adequately counterbalanced. Results of this mixed-design analysis are presented in Table I. The *F* for the two tests is significant at the .001 level; all other "within" subjects effects are not significant. The hypothesis of the study is therefore clearly supported.

Conclusion and Discussion

Adolescent girls, comparable to those used in this sample, express a greater amount of liking for other girls who have previously shown considerable liking of them for their preferences than they express for girls who have previously shown considerable disliking of them on this same basis. This finding, however, must be viewed as something more than a simple exchange of social reinforcements. It was hypothesized in this study that this expression of "like" is verbal behavior which previously has been conditioned to occur in the presence of an approving audience. A test of such an hypothesis will require the use of much younger children as subjects.

Furthermore, it is probable that the subject's expression of "like" was controlled by intraverbal associations, or, to put it another way, was a result of mediated generalization. For example, during the first series most subjects may have made, in spite of the experimenter's effort to discourage them by his instructions, certain covert verbal responses, e.g., "The girl on my left has my preferences." This verbal response may have controlled the expression of "like" if, in the past, subjects have been more frequently reinforced for expressing "like" to listeners whose preferences were similar to their own. Other intraverbal associations may have been acquired by various methods more fully discussed by Skinner (6).

The use of measures of central tendency in this investigation probably has hidden reliable individual differences among these girls. For example, many girls gave more "likes" to the girl who previously disliked them than to the girl who previously liked them. Although in many cases this was probably a matter of chance, in some it may have represented a reliable pattern of behavior; these latter subjects may have "tried to be extra nice" to the disapproving girl. As was suggested parenthetically in the introduction, some persons may have learned to express "like" to a disapproving listener since this has frequently resulted in reducing the disapproval. The reinforcements for such behavior may have also come from "polite" parents.

It has been noted that the present findings are probably a reflection of the prior reinforcement histories of these subjects. Since reinforcements for such verbal behavior as "like" and "dislike" are probably a function of the culture from which the subjects come, it is pertinent to note that the subjects were all above average in intelligence and came from a middle-class suburban area. If it is true that their parents and peer groups place greater

importance upon "being nice to everyone" than do individuals from other social classes, with a different sample the effect observed in this study might be considerably larger.

References

1. Anastasi, A., & Miller, S. Adolescent "prestige factors" in relation to scholastic and socioeconomic variables. *J. soc. Psychol.*, 1949, 29, 43-50.
2. Austin, M. C., & Thompson, G. C. Children's friendships: a study of the bases on which children select and reject their best friends. *J. educ. Psychol.*, 1948, 39, 101-116.
3. Edwards, A. *Personal preference schedule.* Psychological Corporation, 1954.
4. Kuhlen, R. G., & Lee, B. Personality characteristics and social acceptability in adolescence. *J. educ. Psychol.*, 1943, 34, 321-340.
5. Sidowski, J. B., Wyckoff, L., & Tabory, L. The influence of reinforcement and punishment in a minimal social situation. *J. abnorm. soc. Psychol.*, 1956, 52, 115-119.
6. Skinner, B. F. *Verbal behavior.* Appleton-Century-Crofts, 1957.

EFFECT OF WITHDRAWAL OF POSITIVE REINFORCEMENT ON AN EXTINGUISHING RESPONSE IN YOUNG CHILDREN[1]

DONALD M. BAER*
University of Washington

Punishment may be defined systematically in either of two ways: (a) as the presentation of a negative reinforcer for a response or (b) as the removal of a positive reinforcer for a response (9). The first of these methods has seen extensive study in lower organisms (e.g., 4, 8). The second has been subject to a certain amount of discussion in psychoanalytic terminology ("loss of love"), but has seen very little experimental investigation in any species. There are only three exceptions apparent to this statement. Ferster (5) has used a withdrawal procedure ("time out") in studies with lower animals; Brackbill and O'Hare (3) have used it as a punishment technique in facilitating discrimination in young children; and Lewis (6) has used it to set up schedules of reinforcement in young children.

Yet the withdrawal of positive reinforcement as a punishing technique could prove a very significant process in the reinforcement history of the child. Punishment through the presentation of negative reinforcement is typically claimed to be ineffective in weakening responses. Estes concluded that ". . . a response cannot be eliminated from an organism's repertoire more rapidly with the aid of punishment than without it" (4, p. 37). Sears *et al.* claim that punitiveness in a parent ". . . is ineffectual over the long term as a technique for eliminating the kind of behavior toward which it is directed" (7, p. 484). But they also point out that half of the parents they interviewed use punishment at least moderately often and claim that spanking does some good. In the case of the child, it will often prove impossible to present a negative reinforcer without simultaneously withdrawing positive reinforcers. The parent who slaps or spanks almost inevitably will withdraw approval and the more obvious tokens of affection at the same time. Perhaps it is this withdrawal of reinforcement which is the significant technique here.

---

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] The research reported herein was supported in whole by Public Health Service grant M-2208 (C1), United States Department of Health, Education, and Welfare, to whom the author is indebted. Three other debts must be acknowledged: to Mrs. Josephine Reed, director of the Seattle Day Nursery Association Schools, for gracious and generous cooperation; to Mrs. Anne Pilisdorf for her reliable and intelligent performance as *A*; and to Dr. Sidney Bijou for his continuing support, encouragement, and advice.

Child Develpm., 1961, 32, 67-74.

In any case, experimental studies of the punishing effect of the withdrawal of positive reinforcement in the child should prove of systematic and practical value. The present study was designed to yield data on the effectiveness of reinforcement withdrawal for the sole purpose of weakening a response in a situation where a competitive response is not deliberately strengthened through positive reinforcement.

Method

Subjects

Children from a day care center of the Seattle Day Nursery Association were used. All *S*s were experimentally naive. At the outset of the experiment, 16 *S*s were divided randomly into two groups, Punishment and Control. During the course of the experiment, which required that each child participate on five different sessions, six *S*s were lost, usually to their family vacation, occasionally to illness, and in one case, to withdrawal of the child from the school. There remained 10 *S*s, five in each group. The average age of the Punishment group was 6-10, and of the Control group, 6-6. All *S*s were within the normal range of intelligence and represented lower income groups.

Apparatus

The entire experiment was conducted in a mobile laboratory built into a 19-foot house trailer (2), which was parked inside the play yard of the school. The interior of the laboratory included a one-way observation booth for the experimenter and a playroom for the child. The playroom contained a small chair, two tables holding toys or apparatus, a screen on which could be projected cartoons, and a partitioned corner in which an accompanying adult (*A*) could sit, out of the child's sight but still present.

The experiment centered about three pieces of apparatus. The first of these was a translucent screen, on which the experimenter could project cartoons. The screen was mounted in the wall separating child and experimenter; the movie projecter, a Busch "Cinesalesman," operated from the experimenter's side. The projector contained three cartoons on an "endless" reel of film and could present these in an uninterrupted sequence for an indefinite number of cycles. The cartoons were of the Castle Films Woody Woodpecker series (Woody Plays Santa Claus, The Hollywood Matador, and The Dizzy Acrobats); each was black and white, with sound, and lasted seven minutes.

A second piece of apparatus was a white box, approximately one foot on a side, mounted on one of the tables such that *S* could view cartoons and play with this "toy" at the same time. A bar projected from the front surface of the box. At the lower right side of the front surface was a slot from which peanuts could roll as a reinforcement for pressing the bar. Every bar press

resulted in a mild buzz which lasted one second, but did not necessarily produce a peanut. A bar press could also result in a two-second interruption of any cartoon being viewed at the time, if the experimenter chose.

The third piece of apparatus was a commercial toy (Playskool) which was mounted on the other table. It consisted of a cage of sticks, inside of which a small ball could be made to jump up and down a devious route by flipping a small lever at the base of the toy. This toy was not within reach of *S* as he sat viewing cartoon.

Procedure

Ss were dealt with entirely by a young female adult, *A*. Before the experiment started, *A* had been a constant member of the play group and was thoroughly familiar with the *Ss*. Furthermore, she had told them that a trailer was coming and that all the *Ss* would be allowed to enter and play games. This generated a great deal of enthusiasm.

Each *S* was brought to the laboratory on five separate occasions by *A*. The first of these was to adapt the *S* to the laboratory and acquaint him with the cartoons. *A* would seat the *S* in the correct place and retire to her corner. Then the cartoons would start, and all three would play without interruption. At the end of the third cartoon (i.e., after 21 minutes), the projector was stopped, and *A* came out of her corner and took *S* back to the play yard.

The second visit was to establish a learned response in each *S*. Prior to this visit, *A* would ask *S* if he would like to do something different that day. When *S* entered the laboratory, *A* then showed him the toy which could dispense peanuts, saying, "This is a peanut machine." The *S* was seated by it (the same seat as visit 1), and *A* retired to her corner. In very short order, *S* would press the bar and receive a peanut. This started a series of bar presses, of which numbers 1, 2, 3, 5, 7, 10, 15, and 20 were reinforced with a peanut. Immediately after the twentieth response, as *S* was eating the peanut, *A* came out of her corner announcing, "That's all for today." They then left together before the child could respond further.

The third visit was to begin extinction of the learned response, punishing it as well in half of the *Ss*. Prior to this visit, *A* asked *S* simply if he would like to come again. Inside the laboratory, *A* seated *S* in the same seat and retired to her corner. The cartoons were started immediately. *S* was free to watch cartoons and to press the bar as he chose. For the five *Ss* of the Punishment group, any bar press during this session would turn off the ongoing cartoons (picture and sound) for two seconds. For the five Control *Ss*, the bar press had no such action. In either case, the bar press produced the usual buzz, but no peanuts at any time. At the end of the third cartoon the projector was turned off, but *A* remained in her corner for at least two minutes, saying nothing. This gave *S* an interval during which the bar might be pressed, and there were clearly no cartoons to be turned off. Almost every *S* requested to leave by two minutes after the end of the cartoons. In the case of early requests, *A* was instructed to delay until the two minutes

had passed to give each child equal opportunity with the bar. Thus, the visit lasted 23 minutes.

The fourth visit was to allow *S* to show any spontaneous recovery of the learned response, without punishment. Prior to this visit, *A* asked the child if he would like to see the cartoons again. Inside the laboratory, *S* was seated in the usual seat, *A* retired to her corner, and the cartoon was started. *S* could press the bar if he chose, but on this occasion the bar press did not turn off the cartoon, nor did it result in any peanuts. It did produce a buzz as before. Because it was expected that spontaneous recovery would be slight in either group only one cartoon was shown. Again, *S* was kept for two minutes after the end of the cartoon. Thus, this visit lasted nine minutes.

The fifth visit was to examine the strength of the learned response in a situation where punishment was impossible. Prior to this visit, *S* was asked simply if he wanted to come again. Inside, he was told that there would be no cartoons today, but that he could play with anything he wanted. *A* then retired, leaving *S* with only two objects capable of play: the peanut apparatus and the Playskool toy. The peanut machine buzzed but did nothing else when the bar was pressed. This session lasted five minutes, which is rather long for children thoroughly adapted to the laboratory, with only two simple toys to play with.

In summary, then, *S*s were successively adapted to the laboratory and the cartoons, taught a bar pressing response, allowed to extinguish this response with or without punishment, and allowed to show any spontaneous recovery of this response, in situations where the means of punishment was present and absent. This sequence is summarized in Table 1. The interval between the first and second visits was seven days (plus or minus two);

TABLE 1

SEQUENCE OF EXPERIMENTAL CONDITIONS FOR PUNISHMENT AND CONTROL SUBJECTS

Visit	*Purpose*	*Number of Cartoons*	Consequence of Bar Press	
			Punishment Group	*Control Group*
1	Adaptation	3	bar not present	bar not present
2	Strengthen bar press	none	peanuts	peanuts
3	Extinction —session 1	3	cartoon interruption	none
4	Extinction —session 2	1	none	none
5	Extinction —session 3	none	none	none

between the second and third visit, three days (plus or minus one); between the third and fourth visit, five days (plus or minus one); and between the fourth and fifth visit, three days (plus or minus one). All experimentation was conducted during late afternoon hours.

Results

Recordings were made of the bar pressing response on a Gerbrands Harvard Cumulative recorder. Photographic facsimiles of these response curves are shown in Figure 1. Session 1 in this figure refers to *S*'s third visit to the laboratory, during which the bar pressing response resulted in an interruption of the cartoon for Punishment *S*s and otherwise represents an extinction period for all *S*s. Session 2 refers to S's fourth visit, during which one cartoon was seen without punishment. Session 3 refers to *S*'s last visit, during which he played freely with either the bar or the Playskool toy and no cartoons were shown.

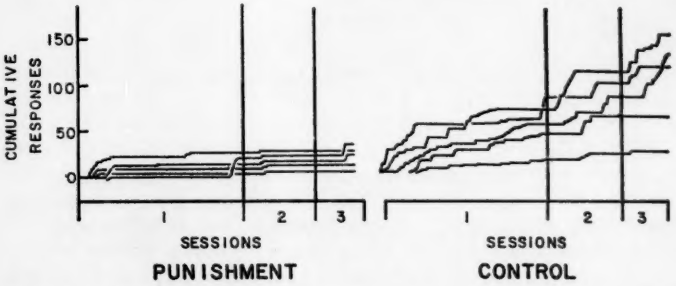

Figure 1—Individual cumulative response curves of Punishment and Control *S*s during repeated sessions.

It is apparent from Figure 1 that the interruption of the cartoon serves to reduce sharply the number of extinction responses *S*s make to the bar. Furthermore, this reduction seems reasonably durable. During the last two minutes of session 1, Punishment *S*s show no particular tendency to increase responding, although there are no cartoons present, and hence they cannot be punished by their interruption. During session 2, where Control *S*s show considerable spontaneous recovery, Punishment *S*s make either one or two responses (which are unpunished) and then cease responding, nor do they respond during the two minutes after the cartoon ends. During session 3, Punishment *S*s direct their play to the Playskool toy rather than to the bar far more than do Control *S*s.

These individual response curves are summarized in Figure 2, which shows the mean number of bar presses made by each group during each

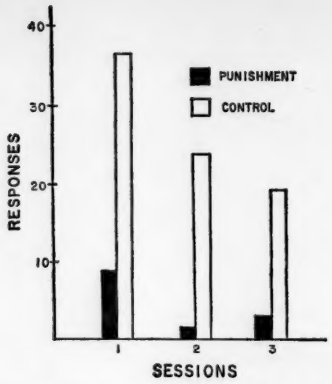

FIGURE 2—Mean responses of Punishment and Control *Ss* during repeated sessions.

session. Essentially the same picture emerges from a consideration of grouped data as from examination of individual curves.

The individual curves present a very clear case. However, there is one case of overlap between the two groups, and because of this an analysis of variance was performed on the data. The Punishment and Control *Ss* constitute two independent groups, and sessions 1, 2, and 3 represent repeated observations of these groups. Use of number of bar presses per session as a raw score is not justifiable because of the heterogeneity of variance between the two groups evident in Figure 1. Consequently, a square root transformation was applied. The transformed data, subjected to Bartlett's test, yielded

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF RESPONSES DURING REPEATED SESSIONS

Source	*df*	*Mean Square*	*F*	*p*
Subjects	9			
Punishment	1	71.10	11.83	.01
Error	8	6.01		
Within Subjects	20			
Sessions	2	11.00	10.78	.01
Sessions × Punishment	2	1.24		
Error	16	1.02		
Total	29			

a χ^2 of 5.28 ($df = 5$, $.30 < p < .50$). A summary of the analysis of variance of the transformed data is given in Table 2. Punishment and Control *Ss* are clearly distinct at the .01 level. The Sessions main effect is of no great consequence, since response is bound to decrease during repeated extinction sessions, especially when the sessions are of successively shorter duration. The absence of any interaction between the Punishment groups and Sessions makes for a gratifyingly simple conclusion.

None of the Punishment *Ss* showed any obvious signs of emotionality during punishment. One *S* in this group, after making several (punished) responses during session 1, was seen to very carefully take hold of the bar, slowly press it down, and, as the cartoon went off, whip his hand away from the bar, saying, "Wow!" Apart from this, all *Ss* took the experiment with great equanimity.

Discussion

From the results obtained, it seems clear that the interruption of a cartoon which the child is viewing will serve very effectively as a punishing event: it will efficiently depress a recently strengthened response in the child's repertoire. The thesis that withdrawal of positive reinforcement is punishing is thereby strengthened, specifically for the case of the child. And it is tempting to wonder if the withdrawal of positive reinforcers may not possess punishing characteristics qualitatively different from the presentation of negative reinforcers.

An essential assumption here is that the cartoons are indeed positively reinforcing stimulation. Other research by the author with the same three cartoons indicated that they have a high and very uniform reinforcing value for children in this age range (1). Testimony to this assumption is that almost any child will learn quickly to press a bar to turn the cartoons back on when they are turned off by the experimenter every few seconds.

It may be important to stress certain aspects of this experiment which could prove critical to its conclusions. One of these is the fact that the response weakened in this procedure is a weak one of short standing. Eight peanuts cannot be taken as a significant amount of reinforcement to a healthy child (who has had juice and cookies in school the same afternoon). These reinforcements, with the increasing ratio schedule on which they are delivered, are sufficient to set up a response strong enough to be studied and manipulated, clearly, but may not give the response much resemblance to other behaviors of the child with more powerful and extensive histories of reinforcement.

Furthermore, in this study punishment has been effected in a very precise manner. Reinforcement was withdrawn immediately consequent to a response in a very consistent way. The fact that this procedure is effective may not guarantee that in a more typical situation, where punishment is offered to a child late, inconsistently, and perhaps incomprehensibly, the effect would be the same.

Finally, the effect of punishment in this study seems durable in terms of the duration of the study. But this is not a great deal of time. Possibly, given more time in the situation, Punishment *Ss* would emit responses equal in number to those of Control *Ss*. It seems more likely, though, that the *Ss* would satiate with the entire situation before this would happen.

Apart from these qualifications to any widespread generalizations, the data give a clear picture of stimulus control in depressing a response in young children. The mechanism is best stated as the withdrawal of positive reinforcement. The technique used here would seem to be a fruitful one for further study of this problem in the laboratory.

Summary

Young children were taught to press a bar for peanuts. During later extinction of this response, five Punishment *Ss* were punished by turning off a cartoon they were watching at the time for two seconds as a consequence of every response. This served to depress the response considerably, relative to five Control *Ss* who were not punished for responding while watching the same cartoons. During later sessions, Control *Ss* showed considerable spontaneous recovery of the response, and Punishment *Ss* did not, even though no longer punished. It was concluded that the withdrawal of positive reinforcement is an effective technique of punishment, at least in situations like the experimental one.

References

1. Baer, D. M. Escape and avoidance response of preschool children to two schedules of reinforcement withdrawal. *J. exp. anal. Behav.*, 1960, 3, 155-159.
2. Bijou, S. W. A child study laboratory on wheels. *Child Develpm.*, 1958, 29, 425-427.
3. Brackbill, Y., & O'Hare, J. Discrimination learning in children as a function of reward and punishment. Paper read at Western Psychol. Ass., Eugene, Oregon, 1957.
4. Estes, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, 1-40.
5. Ferster, C. B. Control of behavior in chimpanzees and pigeons by time out from positive reinforcement. *Psychol. Monogr.*, 1958, 72, 1-38.
6. Lewis, D. J. Partial reinforcement in a gambling situation. *J. exp. Psychol.*, 1952, 43, 447-450.
7. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* Row, Peterson, 1957.
8. Skinner, B. F. *The behavior of organisms.* Appleton-Century-Crofts, 1938.
9. Skinner, B. F. *Science and human behavior.* Macmillan, 1953.

AN EXPLORATION OF PATTERNS OF IMPULSE CONTROL IN EARLY ADOLESCENCE[1]

NORMAN LIVSON *and* WANDA C. BRONSON*
University of California, Berkeley

Somewhere near the core of any current personality theory which lays claim to comprehensiveness there will be found some construct dealing with the fact that most adults are able to control (more or less) and to modify (in any number of ways) the direct expression of their biological impulses. In exploring the development of this phenomenon, some investigators feel more comfortable working within a psychoanalytic framework (e.g., 11); others find learning theory schema more congenial (e.g., 13) or seek to exploit the suggestions provided by a topological model (e.g., 4). We may prefer to speak of "ego development" under the constraints of the "reality principle" or of "socialization" achieved through "learned inhibitions" or of "differentiation of the internal need system" in response to forces in the "life-space." Putting theoretical and stylistic qualms aside, we can probably agree that ". . . by gaining control over the demands of the instincts, by deciding whether they shall be allowed to obtain satisfactions, (and) by postponing satisfaction to times and circumstances favorable in the external world . . ." (5, p. 15), childhood impulsivity becomes gradually transformed into adult responsibility.

While anthropological and psychoanalytic observations and theory regarding the development of impulse control have attended to the full span from infancy to adulthood, empirical investigations leave a gap in the years between childhood and late adolescence. If, as is generally maintained, early adolescence represents a transitional stage, a period of upsurgence of biological impulses, then the modes for handling tensions which the child has already developed would tend, at this time, to assume an increasingly influential role in his psychological functioning. Assuming this, the intent of this study is to explore the linkages of the ways in which early adolescent boys and girls handle their impulses with other aspects of their behavior and personality organization and, further, to examine these relations in the context of the specific values of their possibly very special "subculture." In confining this investigation to the period of early adolescence, we leave aside issues concerning both the genesis of ego control patterns and their

* Guidance Study, Institute of Human Development, University of California, Berkeley 4.

[1] We are pleased to acknowledge the contributions to this report of the staff, past and present, of the Guidance Study and of its director, Dr. Jean W. Macfarlane. We are particularly indebted to Erik Homburger Erikson for his case evaluations, to Dr. Betty Kalis for working out a method for quantification of this material, and to Dr. Mary Darby Rauch for help in carrying out quantification procedures. Drs. Marjorie Honzik and Jack Block have offered helpful criticism of the manuscript.

Child Develpm., 1961, 32, 75-88.

stability over time under the pressures of biological maturation and changes in psychosocial environment.

Method

The source data for this study are comprehensive clinical descriptions of a group of children, prepared in their early adolescence (ages 11 to 13) by E. H. Erikson, after a thorough review and evaluation of all material available from the Guidance Study at that time. The sample employed for this report consists of the 40 children (20 boys, 20 girls) for whom such summaries were completed and represents a rather informal selection from a much larger pool of Guidance Study developmental case histories. The Guidance Study itself is a longitudinal investigation of physical, mental, and personality development in a group of over 200 children who are representative of all children born in Berkeley, California in the years 1928-1929. [For a full presentation of this research program *see* (9).]

Erikson's personality descriptions were made amenable to quantitative analysis through a Q-sorting procedure (8) applied independently by two clinical psychologists to each summary, which yielded composite (summed) item placement scores. From the initial set of 102 Q-sort items, a selection was made of the most highly reliable and discriminating items. The resultant pool of 32 items, and their interrater agreement, is presented in Tables 1 and 2.

Results

Bivariate plots of two Q-sort items, "is undercontrolling; acts upon impulse" and "is rigid, overcontrolling, inflexible," are presented in Figure 1 for boys and for girls.[2] The intercorrelations (*gamma* coefficient[3]) are, respectively, —.44 and —.53. Rather than combining these two measures to obtain a single score for degree of impulse control, each individual was assigned to an ego control group on the basis of his two-variable pattern. This procedure allows for the possibility that such a two-dimensional ego control pattern might provide us with an increment of reliable variance beyond that which could have been carried by some type of single composite score.

---

[2] These items are two of many which were adopted into the present Q-sort pool by Kalis (8) from an earlier version of the California Q-set (2). Although Block and his colleagues are explicit in their conceptualization of "ego control" as a single continuum [cf. (3)], they require separate assessments of degree of "overcontrol" and "undercontrol" in order to permit detection of "bimodal controllers" (individuals who exhibit characteristics of both ends of the dimension) (1). Our use of these two items is a matter of convenience and implies no allegiance to the theoretical model from which they derive; furthermore, the employment of the term "ego control" in this report is casual and intends no distinction from such synonymous (for us) terms as impulse control.

[3] The *gamma* correlation coefficient (6), a variant of *tau*, expresses the degree of covariation between two sets of scores, each of which represents measurement on at least an ordinal scale. Order comparisons of pairs of scores where one or both comparisons involve ties do not affect the value of the coefficient.

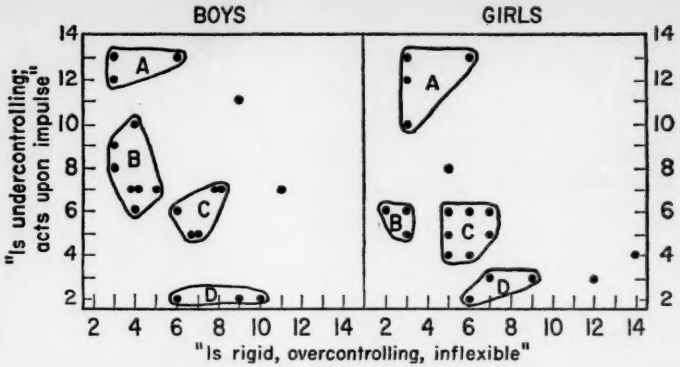

FIGURE 1—Bivariate distributions of ego control scores for boys and girls. The boundaries of the ego control groups, which were determined by visual inspection, are indicated on the scatterplots.

The groupings indicated on the scatterplots were determined by visual inspection with the requirement of at least three cases to form a group. While admittedly failing to insure an objective and unequivocal assignment of cases, this procedure is unbiased: no additional data were consulted and we had no theoretical expectations as to meaningful ego control configurations. The validity of such groupings will be indicated by their ability to generate meaningful distinctions in other aspects of personality organization and functioning.

Derivation of Cluster Scores

Since an examination of the content of the remaining 30 Q-sort items suggested that some reduction to more general dimensions might be possible, separate cluster analyses of boys' and girls' interitem correlation matrices were carried out. Trial clusters were established for each sex on the criterion that each item in a cluster shows a median *gamma* coefficient of at least .40 (approximately at the .05 level[4]) with the other items in that cluster. The results of these analyses, which permitted items to appear in more than one cluster, were examined to determine which of the trial clusters were found for both boys and girls. Only those items which showed similar patterns of interrelations for both sexes were submitted to a new cluster analysis, this time employing interitem correlations computed for the total sample. Final common clusters were established on the following criteria: (a) each item shows a median correlation of at least .50 (approximately $p = .001$) with the other cluster members and (b) this median value is higher with

[4] The significance level of the *gamma* coefficient can be estimated from current work on its sampling distribution carried on by Rosenthal (12).

TABLE 1

RELIABILITIES AND MEDIAN PLACEMENTS OF CLUSTER ITEMS FOR EGO CONTROL GROUPS

Item	Reliability†	Boys A N=3	Boys B N=7	Boys C N=5	Boys D N=3	Girls A N=4	Girls B N=3	Girls C N=7	Girls D N=3
Ego Control									
2. Is undercontrolling; acts upon impulse ..	.62	13	7	6	2	12.5	6	5	3
6. Is rigid, overcontrolling, inflexible	.57	3	4	7	9	3.0	3	6	7
Ego Strength Cluster									
5. Tends to make personally relevant many different areas; to become ego-involved ..	.59	6	9	5	6	7.0	9	11	7
11. Is relaxed, easy-going	.74	3	10	5	5	4.0	7	10	6
13. Absorbs family's cultural values without strain	.47	5	11	6	6	6.0	5	11	10
4.* (Does not) lack organization, synthesizing of ego strengths .	.35	6	12	6	8	4.5	10	10	9
7.* (Does not) have poor controls for anxiety .	.71	3	11	5	8	4.0	10	12	11
Social Adjustment Cluster									
1. Is self-confident and assured in relations with others	.70	8	11	4	3	5.5	8	10	10
16. Is accepted as a leader by peers	.62	7	8	7	6	5.0	9	9	10
17. Is popular with other children	.82	8	8	7	7	4.5	10	11	9
18. Has good role variability	.64	6	10	4	3	4.5	11	12	6
20.* Is (not) awkward, clumsy in social situations	.70	7	10	6	8	6.5	11	12	13
22.* Is (not) socially constricted; (does not have) a narrow range of social behavior ...	.75	9	11	5	5	7.0	12	12	8
Emotional Expansiveness Cluster									
19. Tends to be outgoing in social situations ..	.59	11	11	7	4	11.0	11	12	8
21.* Is (not) aloof, distant in personal interactions	.75	11	7	4	6	10.5	8	8	6
24.* Is (not) bland, emotionless in approach to situations	.61	12	10	6	5	11.5	10	11	8
32.* Is (not) emotionally constricted	.52	10	9	4	3	8.5	10	11	4

* Indicates "negative" cluster item for which scale value is reversed. Material in parentheses brings the item into proper orientation with the cluster.

† Agreement between two sorters evaluated by the *gamma* correlation coefficient (6).

its own than with any of the other clusters. Therefore, in this final clustering, no item can appear in more than a single cluster. Also, at least three items are required to define a common cluster. This procedure yielded three clusters which seemed readily identifiable as Ego Strength (five items), Social Adjustment (six items), and Emotional Expansiveness (four items). The item composition of the clusters is presented in Table 1, and a listing of the 15 residual items which failed to meet the clustering criteria may be found in Table 2.

Each child's score on a given cluster is his mean item placement over all items comprising the cluster, the scale values for "negative" items being reversed. Since item scores are the sum of placements by two sorters on a seven-category Q sort, the resultant values for both cluster and individual item scores may range from 2 to 14, higher scores denoting a greater amount of the given quality.

Intercluster (*gamma*) correlations based on these scores are as follows: Ego Strength × Social Adjustment (boys, .57; girls, .77); Ego Strength × Emotional Expansiveness (boys, .36; girls, .31); Social Adjustment × Emotional Expansiveness (boys, .46; girls, .40). For both sexes, the interrelations among clusters are quite substantial (especially when the reliabilities of the component items are considered) so that a general factor, perhaps loosely defined as "psychological health," might be substituted for the three separate clusters. However, since the items within each cluster could be sensibly assigned to conceptually independent dimensions, it seemed legitimate—and possibly useful—to keep them distinct.

Characteristics of the Ego Control Groups

In this attempt to define the correlates of the different ego control patterns we shall rely primarily upon group differences in the three cluster scores. However, a reading of the median group values for the unclustered (residual) items (Table 2) will be employed to evaluate and supplement the resultant characterization of the ego control groups. Two criteria are used in reporting discriminating residual items: (a) all extreme items (scale values above 10 or below 6) are cited as "descriptive" of a given group and (b) statements of intergroup "differences" require at least a two-point scale difference. The significance levels of intergroup differences (cited in Table 2) should be regarded as only rough guides to their relative reliability, due to the low and highly variable power of statistical evaluations with such small groups.

Boys. Figure 2 (left) summarizes the median cluster scores for the four boys' ego control groups. (The median values for individual items within each cluster are presented in Table 1.)

Group A, the most undercontrolled and the most not overcontrolled pattern, shows lowest Ego Strength, moderate Social Adjustment, and (not surprisingly in light of its definition) the most extreme degree of Emotional Expansiveness. The syndrome becomes more recognizable when we con-

TABLE 2

RELIABILITIES AND MEDIAN PLACEMENTS OF RESIDUAL ITEMS FOR EGO CONTROL GROUPS

Item	Reliability†	Boys A N=3	Boys B N=7	Boys C N=5	Boys D N=3	Boys Diff.*	Girls A N=4	Girls B N=3	Girls C N=7	Girls D N=3	Girls Diff.*
3. Is independent in thoughts and feelings	.55	5	9	4	7	A<B .12 B>C .02 C<D .15	5.5	4	9	10	.....
8. Shows conflict about sexual identification	.48	11	6	12	11	B<C .05	11.0	7	9	11	A>B .06
9. Identifies with mother or mother-figures	.69	9	7	10	7	C>D .06	12.0	10	12	10	.....
10. Identifies with father or father-figures	.53	9	10	8	11	C<D .04	8.5	5	7	7	.....
12. Represses unacceptable impulses	.52	8	9	9	9	.....	6.0	12	9	10	A<B .06 A<C .03 A<D .06 B>C .03
14. Is motivated in behavior by guilt feelings	.79	8	5	8	6	B<C .15	6.5	8	7	7	.....
15. Is individualistic, nonconforming	.40	8	5	5	8	A>C .08	10.0	3	4	4	A>B .06 A>C .05 A>D .09
23. Is openly aggressive, hostile in relations with others	.44	11	4	6	6	.....	9.5	5	5	5	A>B .09 A>C .04 A>D .06
25. Is a verbally expressive individual	.49	8	8	6	11	.....	9.5	10	10	8	.....

(*table continued on next page*)

TABLE 2 (*continued*)

RELIABILITIES AND MEDIAN PLACEMENTS OF RESIDUAL ITEMS FOR EGO CONTROL GROUPS

Item	Reliability†	Boys A N=3	Boys B N=7	Boys C N=5	Boys D N=3	Boys Diff.*	Girls A N=4	Girls B N=3	Girls C N=7	Girls D N=3	Girls Diff.*
26. Has passive approach to environment	.73	4	4	12	9	A<C .04 A<D .10 B<C .11	2.5	4	5	3	A<C .14
27. Seeks attention and praise	.67	13	10	10	12	A>C .12	13.0	10	9	11	A>B .09 A>C .05
28. Tends to take roles of mothering as reflection of nurturant needs	.66	7	8	6	7	A>C .15 B>C .07 B>D .12	10.0	12	8	13	C<D .10
29. Is unduly fearless, counterphobic	.60	12	5	6	12	A>B .04 A>C .04 B<D .12 C<D .15	10.0	5	6	4	A>B .09 A>D .06
30. Is dominating, assertive	.43	10	6	7	7	A>C .04 A>D .10	10.5	7	11	8	.....
31. Is timid, fearful	.35	6	8	11	7	A<C .04 A<D .10 B<C .10 C>D .04	5.5	10	6	8	A<B .12 B>C .12

* Group differences evaluated by Mann-Whitney U test ($p \leq .15$ reported).

† Agreement between two sorters evaluated by *gamma* correlation coefficient.

sider the tendency of these boys to be fearless and counterphobic (item 29), hostile (item 23), and dominating (item 30), and not passive (item 26). However, they lack independence (item 3) and require a great deal of attention and praise (item 27). Thus, while their undeniable assertiveness may earn some degree of social success within the peer group, it appears to serve the defensive function of masking an underlying lack of security.

Group B, while also not overcontrolled but appreciably less impulsive than group A, presents a cluster profile suggestive of a uniquely high level of psychological health: these children score high, both relatively and absolutely, on Ego Strength and Social Adjustment and, while also scoring relatively high on Emotional Expansiveness, they show somewhat more restraint than the group A boys. This impression is enhanced by the residual items: of all groups, group B boys are the most independent in thoughts and feelings (item 3), show least conflict over sexual identification (item 8), and are not motivated by guilt (item 14). Further, they are neither counterphobically fearless (item 29) nor hostile or aggressive (item 23), but any suggestion of general submissiveness seems denied by their lack of passivity (item 26).

Groups C and D are indistinguishable with respect to their cluster scores: both have very low Ego Strength (scoring as low as group A), are considerably lower than either group A or group B on Social Adjustment, and show strikingly lower Emotional Expansiveness. However, despite this similarity, some differentiation between these two groups is suggested by residual items. While boys of both groups are far more passive in their approach to the environment (item 26) than any of the other boys, group C is more passive than group D and also more timid and fearful (item 31). Group D boys, on the other hand, show greater independence of thoughts and feelings (item 3), are more individualistic (item 15), seek more attention and praise (item 27), and are much more counterphobically fearless (item 29). Further, they are quite verbally expressive (item 25), much more so than any of the other groups. This, when combined with the finding that the mean mental test scores (Stanford-Binet, ages 6-15) of the three group D boys are among the four highest found in our total Guidance Study sample, suggests that they may qualify as the "brains" of their peer group. Groups C and D also exhibit different identification patterns: group C boys are more identified with their mothers (item 9); group D boys, with their fathers (item 10). Furthermore, only group C, among all groups, shows stronger identification with the mother than with the father. Taken together, these data suggest that, despite their similarity on cluster scores, the two groups may present different personality structures. Lack of emotional expansiveness, for example, may be for group C a direct expression of underlying passivity; for group D, it may reflect defensive reserve.

Girls. Figure 2 summarizes the median cluster scores for the four girls' ego control groups. (Again the median values for individual items in each cluster can be found in Table 1.)

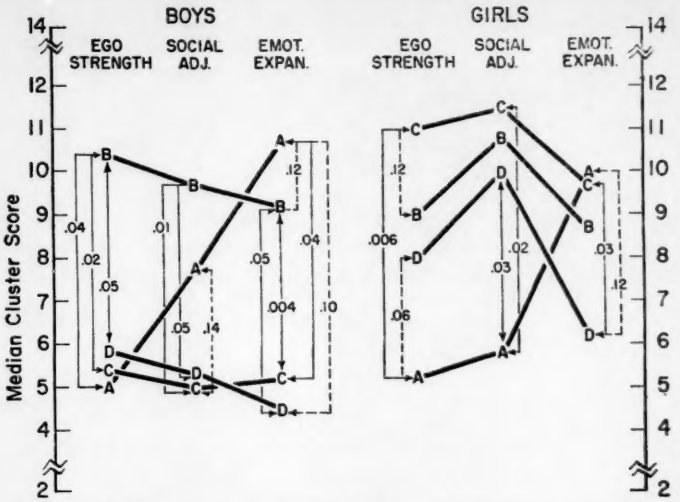

FIGURE 2—Median cluster scores for boys' and girls' ego control groups. Differences between groups significant at or below the .05 level are indicated by solid lines; probability levels between .05 and .15, by broken lines.

The girls' group A presents an undercontrolled pattern as pronounced as that of the boys'. Their low Ego Strength and high Emotional Expansiveness scores are strikingly similar to the boys' pattern, but, in sharp contrast, they score extremely low on Social Adjustment. An indication of the nature of their social behavior is provided by the residual items: while all other girls are uniformly low on aggressiveness (item 23) and nonconformity (item 15), group A girls characteristically score high on these two items. Furthermore, they are fearless (item 29), dominating (item 30), and not timid (item 31), although they lack independence of thoughts and feelings (item 3) and seek attention and praise (item 27). Consonant with the "acting-out" behavior shown by group A girls is their clear differentiation from other girls in showing less tendency to repress unacceptable impulses (item 12). Lastly, although these girls are strongly identified with their mothers (item 9), they show considerable conflict over their sexual identification (item 8).

Only group A girls show a distinctive cluster score profile; the remaining groups appear very similar, on residual items as well as on cluster scores. Despite this apparent homogeneity, it is possible to assign distinguishing characteristics to each of the groups, aided, in this instance, by an examination of *intra*cluster item differences among them.

Group C girls, who tend to be neither impulsive nor overcontrolled, show the highest degree of Ego Strength and Social Adjustment and are quite high on Emotional Expansiveness. On residual items, these girls are the most assertive (item 30) and the least timid (item 31) of the three groups. They differ further in being only moderately nurturant (item 28) —in contrast with the extremely nurturant group B and group D girls— while showing stronger identification with their mothers (item 9) than either of these groups. An analysis of the component items of the Ego Strength cluster demonstrates that, although they score uniformly high on each, their advantage over groups B and D derives primarily from three items: they appear more relaxed and easy going (item 11), they become ego involved in a variety of areas (item 5), and they are able to absorb their families' values without undue strain (item 13). The impression we gain, then, is that group C girls are characterized by considerable personal stability and interpersonal ease.

Group B girls are intermediate between groups C and D on all three of the cluster variables. Of the few characteristics which do distinguish them from the C and D groups, the clearest is their lack of independence of thoughts and feelings (item 3). Further, they are more repressive (item 12) and timid (item 31) and, although they share with groups C and D girls an extreme tendency toward conformity (item 15), group B girls experience difficulty in absorbing their families' cultural values without strain (item 13, Ego Strength cluster). Of all the girls' groups, they show the least identification with the father (item 10).

Only group D girls are clearly not undercontrolled; they also achieve the highest, although still only moderate, degree of overcontrol. They are best separated from all other groups by their very low Emotional Expansiveness. On both Ego Strength and Social Adjustment, group D girls are lower than those of groups B and C but still fall well within the middle range of the scales. The residual items permit little differentiation of these girls from the other groups. Group D girls, like those in groups B and C, are neither passive (item 26), individualistic (item 15), fearless (item 29), nor aggressive (item 23). They do seek attention and praise (item 27) and, in comparison with B and C girls, are somewhat less verbal (item 25). These characteristics, combined with the apparent stereotypy of their social behavior (items 18 and 22, Social Adjustment cluster) and their emotional constriction (item 32, Emotional Expansiveness cluster) suggest that group D girls adopt a "little lady" role which may serve a defensive function and, specifically, perhaps may relate to their conflict over sexual identification (item 8).

Discussion

While the personality sketches obtained through the ego control groupings are obviously incomplete and necessarily tentative, their feasibility supports the relevance of this parameter to personality organization in early

adolescence. Furthermore, these syndromes encompass characteristics rather remote from those directly inferable from the ways in which the child manages his impulses during this period. For example, the rather "average" ego control pattern of group B girls cannot anticipate their distinct tendency to employ repressive defenses; the nature and the similarity of group C and group D boys' control scores cannot directly account for their opposite parent identification patterns.

Our use of two dimensions to characterize impulse control has been due as much to expediency as to any prior theoretical notions (*see* footnote 2). We find, however, that the two-variable control groupings have apparently allowed for greater discrimination among individuals than would have been suggested by a frequency distribution of a single measure of impulse control. From such a single score, for example, the distinction between boys' groups B and C, easily made from the scatterplot, would not have arisen and the rather suggestive differences between them would not have been detected. Certainly, it is not difficult—on a descriptive level, at least—to regard impulsivity as implying something more than the polar opposite of inflexibility. When we say of a person that he is "anything but" rigid and overcontrolled, we evoke an image of spontaneity but not necessarily one of extreme impulsivity. Nevertheless, the present results provide insufficient basis for a serious evaluation of the dimensionality of ego control as a theoretical construct; they do, however, indicate the promise of such an undertaking.

Any generalization of the results of this investigation to other developmental periods, or even to early adolescence experienced under other value systems, would seem to be a very risky endeavor. When the peer culture differs, or as it changes with approaching adult status, an adolescent's impulse control pattern may find different expressions and induce different social reactions. For example, boys' and girls' groups A share a pattern which links considerable impulsivity with emotional disturbance and extreme lability of expression. Yet, for boys, this syndrome permits an "average" level of social adjustment while, for girls, the resultant is pronounced social failure. This suggests that the prescriptions for socially appropriate behavior are widely divergent for boys and girls in early adolescent peer culture, specifically with respect to its sanctioning of impulsive behavior.

Unusually convenient support for this hypothesis is provided by Tuddenham (14) who conducted sociometric studies in the classrooms of these same children in the early school grades. He reports that in the fifth grade (approximately age 11) boys who are seen by their peers as "taking chances" also tend to be "popular" with their classmates and to be regarded as "leaders" of the group. Daring is not so rewarded for girls. Thus, the differing social adjustment of undercontrolling boys and girls may be tied to these contrasting peer evaluations, mediated by the differing self-perceptions they induce in the young adolescent. Furthermore, since these data also indicate that impulsivity in boys becomes steadily less valued through the school

years, we may infer that group A boys will come under pressure to reappraise themselves and their behavior in relation to changing peer mores. Under such influences, the ego control pattern itself may be modified or, as the heterogeneity of peer culture increases with increasing age, the boy may tend to preserve his pattern by a self-selection of a more congenial social milieu which accepts, or even rewards, his impulsivity.

The two facets of impulse control appear to have somewhat different connotations for the two sexes. Examining the correlations between the two ego control items and the cluster scores over the total sample (Table 3), we find that boys' Ego Strength and Social Adjustment bear little relation to their degree of impulsivity but are substantially related to the inflexibility measure. For girls, the reverse conclusion is suggested: for them, degree of overcontrol is irrelevant while effective functioning tends to accompany a lack of impulsivity. A clue to an understanding of these relations is to be found in the interaction of the different social mores and maturational levels of boys and girls at this particular period. We may hypothesize that girls, who are biologically more mature at this age, are required to exert effective control over sexual impulses if they are to maintain satisfactory adjustment to self and to society. For boys, sexual undercontrol is less at issue while, on the other hand, excessive control of aggressive, assertive impulses is a social and personal handicap in early adolescence. From these premises the apparent greater importance of degree of impulsivity for girls, and degree of overcontrol for boys, to their respective personality organization may be deduced.

Apart from sex differences in rate of physical maturation during adolescence, variations within each sex on developmental status may be suspected to be of some relevance to ego control. A number of maturational indices were selected for comparing ego control groups and for correlation with the two control measures: age at maximum growth (MGA) and a "general factor" maturity score (10) for both sexes, age at menarche for girls, and age at reaching Sex Stage IV (7) for boys. For neither sex do the ego control groups differ significantly on any of the maturational indices. For boys, both control scores show near-zero correlations with each of the

TABLE 3

CORRELATIONS OF EGO CONTROL ITEMS WITH CLUSTER SCORES

	Ego Strength		Social Adjustment		Emotional Expansiveness	
Item	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Is undercontrolling, acts upon impulse	—.13	—.24	.26	—.24	.46	.27
Is rigid, overcontrolling, inflexible	—.48	.02	—.57	.06	—.82	—.27

three maturational indices, but, for girls, there is the curious finding that MGA and menarcheal age, despite their strong positive intercorrelation ($r = .71$ for all Guidance Study girls), show substantial and *opposite* correlations with degree of overcontrol (MGA, *gamma* = .40; menarche, *gamma* = −.31). (The same reversal, but with lower values, is found for undercontrol.) Inspection of scatterplots reveals that girls whose growth spurt is relatively early (before age 11) and those whose menarche occurs later than average (after age 13) tend towards extreme overcontrol of impulses. The remainder of the range on both maturational indices shows no relation to either control item. This suggests that the interval between MGA and menarche may relate to the rigidity and inflexibility of impulse control. We find that girls for whom MGA precedes menarche by two or more years have a mean overcontrol score of 12.6 (7.8 on the impulsivity measure); for those who reach both of these developmental points within the same year, the corresponding values are 7.8 and 12.5. For the remaining girls, with the modal interval of one to two years, the mean scores are intermediate: 10.1 and 10.3. While these three groups vary somewhat, and in the obvious directions, on mean age for MGA and menarche, it seems more probable that it is the timing relation between rate of growth in size and sexual maturation which underlies their variation along the ego control dimensions. Why a relatively tall and sexually immature girl should tend toward greater inflexibility in handling her impulses during early adolescence than one who is small and relatively mature—if, indeed, this proves to be a replicable finding—is difficult to understand. What it does argue, however, is that, despite the existence of a strong general factor coordinating the various processes of maturation during adolescence, there may be useful information in the temporal patterning of these processes. If this is so, then —to take one example—investigations of personality correlates of maturational rate may be made more powerful by a consideration of the possible impact of asynchronies among the various components of development in adolescence.

The present analysis of the behavioral and intrapsychic correlates of ego control is thus qualified by its limitation to a given period of development (and the different maturity levels of the two sexes during that period) and to the special qualities of early adolescent culture. Specifying the link between what we have found for this period and the picture which will be discernible in adulthood—whether it will indicate constancy, predictable transformations, or puzzling discontinuities—must await the completion of a current assessment of these persons, now 30 years of age.

Summary

Samples of children in their early adolescence (20 of each sex) were partitioned into eight ego control groups on the basis of clusters found visually in bivariate distributions of ratings of overcontrol and undercontrol of

impulses. Differences among these groups on three primary variables (Ego Strength, Social Adjustment, and Emotional Expansiveness) were found which, when combined with group differences occurring in a supplementary set of personality evaluations, permitted the assignment of partial and tentative personality syndromes to each of the ego control patterns. Other parameters, such as the sex of the child and differential maturational rates, determine the specific natures of the relations between control mode and personality organization.

References

1. Block, J. The development of an MMPI-based scale to measure ego control. Institute of Personality Assessment and Research Report, 1953 (mimeo).
2. Block, J. *The Q-sort method in personality assessment and psychiatric research.* C. C. Thomas, 1961.
3. Block, J., & Thomas, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.,* 1955, 51, 254-259.
4. Block, Jeanne, & Martin, B. Prediction of behavior of children under frustration. *J. abnorm. soc. Psychol.,* 1955, 51, 281-285.
5. Freud, S. *An outline of psychoanalysis.* Norton, 1949.
6. Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications. *J. Amer. statist. Ass.,* 1954, 49, 732-769.
7. Greulich, W. W., *et al.* Somatic and endocrine studies of puberal and adolescent boys. *Monogr. Soc. Res. Child Develpm.,* 1942, 7, No. 3 (Serial No. 33).
8. Kalis, B. L. Testing hypotheses with narrative case material. Paper read at 14th V. A. Regional Psychology Research Conference of the San Francisco Bay Area, Berkeley, July, 1956.
9. Macfarlane, J. W. Studies in child guidance: I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.,* 1938, 3, No. 6 (Serial No. 19).
10. Nicolson, A. B., & Hanley, C. Indices of physiological maturity: derivation and interrelationships. *Child Develpm.,* 1953, 24, 3-38.
11. Redl, F., & Wineman, D. *Children who hate.* Free Press, 1951.
12. Rosenthal, I. Sampling distribution of the nonparametric correlation coefficient, gamma. Paper read at Western Psychol. Ass., San Diego, April, 1959.
13. Sears, R. R., Whiting, J. W. M., Nowlis, V., & Sears, P. S. Some childrearing antecedents of aggression and dependency in growing children. *Genet. Psychol. Monogr.,* 1953, 47, 135-234.
14. Tuddenham, R. D. Studies in reputation: III. Correlates of popularity among elementary school children. *J. educ. Psychol.,* 1951, 42, 257-276.

LEVEL OF STIMULUS PRETRAINING AND PAIRED-ASSOCIATE LEARNING [1]

HAYNE W. REESE*
University of Buffalo

Previous studies of the effects of varying amounts of stimulus pretraining on subsequent discrimination or paired-associate learning have yielded inconclusive results. Some reported significantly more facilitation following greater amounts of pretraining, but did not control for the effects of possible performance set differences in groups given different amounts of pretraining (e.g., 2, 5). Others controlled performance set, but found that different amounts of pretraining did not produce significantly different amounts of performance set or of acquired distinctiveness (1, 3). The *S*s, stimuli, and task I responses were comparable in these studies, but the former studies used button-pushing or switch-throwing responses in task II, whereas the latter studies used the Star Discrimeter. In the present study, stimuli and responses comparable to those used in the former studies were used, but children were used as *S*s because it was desired to simplify the tasks and to minimize the probability that *S*s would have preexperimentally learned names for the stimuli used in the condition which controlled performance set effects.

Two specific predictions were tested: (a) the higher the level of pretraining, the better the performance on task II will be; and (b) the higher the level of pretraining, the greater the difference between the experimental (acquired distinctiveness) and performance-set control conditions will be.

In task I, *S*s learned to associate distinctive nonsense syllables with two stimuli. Three groups were trained to a low, moderate, or high criterion.

* Department of Psychology, University of Buffalo, Buffalo 14, New York.

[1] This research was conducted while the writer was a research assistant at, and with funds made available by, the Iowa Child Welfare Research Station. The writer is indebted to Dr. Charles C. Spiker for his aid in the research.

Child Develpm., 1961, 32, 89-93.

In task II, *Ss* learned to associate button-pushing responses with these two stimuli and two others. The two stimuli used in both tasks constitute the experimental condition, and the two used only in task II constitute the control condition. Presumably, performance set affects performance on the control condition, and both performance set and acquired distinctiveness affect performance on the experimental condition. Therefore, the difference between performance on the two conditions can be attributed to acquired distinctiveness, with performance set effects controlled.

Method

Subjects

The *Ss* were 90 fourth and fifth grade school children from two public schools in Ottumwa, Iowa.[2] They were assigned to the experimental groups in a random manner.

Apparatus

The apparatus was essentially the same as that described in detail by Spiker and Holton (6), except that the response unit contained only two push-buttons, located side-by-side approximately 5 in. apart. There was a small pilot lamp above each button.

The stimulus unit contained a single stimulus aperture through which lights differing in hue could be presented one at a time. Corning Glass Co. filters were used to produce four stimuli: two red lights, a light red (No. 2418) and a dark red (No. 2030); and two green lights, a light green (No. 3718 and No. 4784 together) and a dark green (Nos. 3384, 3389, and 5031 together).

The task I response words were three-letter nonsense syllables, *lev* and *wog*. In terms of common elements and speech sounds the syllables are highly distinctive. The syllables were printed on 3 in. by 5 in. cards, which were presented manually by *E*.

Procedure

Pretraining task. The response unit was concealed throughout task I. The *Ss* learned to associate the nonsense syllables with two stimuli in task I. Group H was trained to a high criterion, 19 correct responses in 20 consecutive stimulus presentations; group M was trained to a moderate criterion, seven correct responses in 10 consecutive presentations; and group L was trained to a low criterion, three consecutive correct responses. The red stimuli were used for half of each group, and the green stimuli for the other half. There were 30 *Ss* in each of the three experimental groups.

[2] The writer is indebted to Mr. R. O. Wright, Director of Curriculum, Ottumwa Public School System, Ottumwa, Iowa, for his generous permission to use the *Ss* of this experiment. Thanks are also due Mr. Glenn J. James, Principal, Hedrick and Fairview Elementary Schools, and his teaching staff, for their assistance in many ways.

The anticipation time was 2 sec., and the end of this interval was signalled by a bell. When *S* gave an incorrect response, or failed to respond, he was shown the card bearing the correct syllable during a 3-sec. joint-presentation interval and was required to read it aloud. The interstimulus interval was 2 sec.

Transfer task. Immediately after *S* reached the appropriate task I learning criterion, the response unit was placed in position in front of the stimulus-exposure unit and below the stimulus aperture. All four stimuli were used in task II; *S*s learned to associate light green and dark red with the left button, and dark green and light red with the right button. The pilot lamp above the correct button came on at the end of a 3-sec. anticipation time and remained on during a 2-sec. joint-presentation interval. The interstimulus interval was 2 sec. The *S*s were instructed to respond prior to the onset of the pilot lamp and were allowed only one response to each stimulus presentation. The bell rang immediately after correct responses.

All *S*s were given 18 trials, each trial consisting of the presentation of all four stimuli once each. No stimulus was repeated consecutively, and no stimulus alternated with one of similar hue more than once.

Results

The mean numbers of stimulus presentations through criterion in task I were: group L, 6.1 (SD = 3.2); group M, 9.3 (SD = 3.3); and group H, 32.3 (SD = 21.5).

The mean numbers of errors in blocks of six trials on task II are given in Table 1. Analysis of variance of these data indicated that the difference between the over-all means for the experimental and control conditions was significant ($F = 4.61$; $df = 1, 87$; $p < .05$). This effect did not interact

Table 1

MEAN NUMBER OF ERRORS ON TASK II

Condition	*Blocks of Six Trials* 1	2	3	*All Trials*
Group L				
Control	5.6	4.1	4.3	4.7
Experimental	5.3	3.9	3.1	4.1
Group M				
Control	5.4	4.2	3.2	4.3
Experimental	4.6	3.5	3.4	3.8
Group H				
Control	5.5	2.7	3.0	3.7
Experimental	3.8	3.1	2.4	3.1

significantly with Levels of Pretraining ($F < 1.00$), but the triple interaction of Conditions, Levels of Pretraining, and Trials was significant ($F = 2.84$; $df = 4, 174$; $p < .025$). The simple effects involved in this interaction were tested by analysis of variance of the data for each block of trials separately. These analyses revealed only one significant effect; in the first six trials the Conditions effect was significant ($F = 9.86$; $df = 1, 87$; $p < .005$), indicating that performance on the experimental condition was significantly superior to that on the control condition only in the first block of trials. Examination of the means on the first block of trials shows that the difference between the control and experimental conditions is much greater in group H than in group L, but, since Conditions did not interact significantly with Level of Pretraining ($F = 1.74$; $df = 2, 87$; $p > .10$), this effect was not statistically reliable.

Discussion

The analyses indicated that the facilitation attributable to acquired distinctiveness was significantly greater than that attributable to performance set only in the first block of trials. This finding would be expected on the assumption that the experimental and control learning curves approach the same asymptote, so that differences become smaller as learning progresses.

Contrary to the second prediction, there was no significant increase in acquired distinctiveness with increased level of pretraining, suggesting that acquired distinctiveness is maximal after very low levels of pretraining, about six stimulus presentations in the present study. None of the theoretical mechanisms which have been suggested account for the rapidity of the development of the effect. Those based on learned mechanisms would seem to require a greater number of pretraining trials to permit learning to occur, whereas those based on perceptual responses do not lead to the prediction of a difference between the experimental and control conditions when these conditions are within-*Ss* factors and the stimuli within each condition differ on the same dimensions, e.g., saturation and brightness in the present study.

The data suggested that performance set may have increased with increasing level of pretraining, and, although the analyses of variance did not support this suggestion, *t* tests of the differences between groups L and H (the extreme groups on the independent variable) offered some support, in that the difference in the second block of trials approached significance ($t = 1.90$; $df = 87$; $p = .06$), and the difference over all trials was significant ($t = 2.06$; $df = 87$; $p < .05$). These *t* tests do not provide statistical justification for an assertion that performance set must be controlled in studies of acquired distinctiveness, especially since previous studies which controlled performance set found that it did not increase significantly (1, 3, 4), but the data of these and the present study suggest that such a trend may be present and that performance set should be controlled.

Summary

It was predicted that (a) the higher the level of stimulus pretraining, the better the performance on a motor paired-associate transfer task will be; and (b) the higher the level of stimulus pretraining, the greater the difference between an acquired distinctiveness condition and a performance set control condition will be.

Ninety fourth and fifth grade school children learned distinctive nonsense-syllable names for two similar stimuli in task I. Group L learned the associations to a low criterion; group M, to an intermediate criterion; and group H, to a high criterion. In task II, *Ss* learned to associate button-pushing responses with these stimuli and two others.

The first prediction was supported only by a test of the difference between the task II performance of groups L and H, which suggested the presence of a trend. The second prediction was not verified; the difference between the experimental and control conditions was significant in the first block of trials, indicating that the facilitation on the experimental condition was attributable not only to performance set but also to acquired distinctiveness, but this effect was not significantly different in the three groups.

References

1. Cantor, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, 50, 180-184.
2. Gagne, R. M., & Baker, K. E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, 40, 439-451.
3. McCormack, P. D. Negative transfer in motor performance following a critical amount of verbal pretraining. *Percept. mot. Skills*, 1958, 8, 27-31.
4. Reese, H. W. Motor paired-associate learning and stimulus pretraining. *Child Develpm.*, 1960, 31, 505-513.
5. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, 42, 173-182.
6. Spiker, C. C., & Holton, R. B. Associative transfer in motor paired-associate learning as a function of amount of first-task practice. *J. exp. Psychol.*, 1958, 56, 123-132.

MEASUREMENT OF MOVEMENT IN THE HUMAN NEWBORN: A NEW TECHNIQUE [1,2]

WILLIAM KESSEN,* LOUISE S. HENDRY, *and*
ANNE-MARIE LEUTZENDORFF
Yale University

The behavior of the human newborn shows little of the intentional character of later infant or adult behavior; that is, the responses of the newborn can only tenuously and with difficulty be related to the events in his environment. The newborn behaves without goals (10). Largely for this reason, no generally accepted taxonomy of responses has been developed for neonatal behavior, and researchers have used a wide range of techniques in their attempts to make sense of the disconcerting variety of neonatal movement. The most highly regarded techniques used in the last several decades have aimed at description of "mass activity" or "general activity" rather than at the description of circumscribed local movements. Even with this limitation on the specification of the response-class to be studied, however, there has been a great deal of diversity in technical methods for recording activity or movement. The human observer has been used, without instrumentation other than pencil and paper, to catch the complexities of neonatal behavior (16), although the more common employment of the gadget-free human observer has been restricted by checklists or rating scales (3, 4, 14). Other than the unassisted human being, the technique of choice in the recording of neonatal movement has been one form or another of the stabilimeter (3, 6, 11, 14). This ingenious device, closely akin to mechanisms used to assess activity in animals (e.g., 2), provides a numerical index of

* Department of Psychology and Child Study Center, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported here was supported in part by Research Grant M-1787, United States Public Health Service.

[2] The writers express their thanks to Dr. Milton J. E. Senn and to the staff of Grace-New Haven Community Hospital for their generous cooperation in support of the research. E. J. Williams and J. P. Williams assisted in the collection of data and their analysis.

the infant's movement, an index which has been shown to be sensitive to environmental variations of several kinds (Irwin, *passim*). The advantages of stabilimetric techniques over the unaided human observer in a quantitative treatment of newborn movement are clear: The danger of observer bias and contamination is reduced markedly; there is no shifting focus on different limbs; the baby's movements can be recorded over long stretches of time; and there is relatively little ambiguity in scoring the data-protocols. This is not to say that the human observer is an obsolete device in the study of neonatal behavior. At current levels of ignorance about what makes the neonatal organism work, it is of crucial importance that the sensitive human instrument be used to develop hunches, to make preliminary confirmations of these hunches in general unstructured observation, and, in the case of well-developed checklists or rating scales, to test hypotheses systematically. Within the limits of these qualifications, objective and reproducible techniques for recording movement will have substantial value for our understanding of the psychology of the neonate.

There has been available for many years a recording technique which would seem at first glance to be ideally suited to a study of neonatal movement—the motion-picture camera. Surprisingly enough, although films of young children have often been made (7, 9, 13), these films have only rarely been put to quantitative analysis (17). Gesell and Halverson (8) made systematic cinematographic observation of a single child over a period of several months, scoring for posture and laterality. Gilmer (9) has used film to describe in greater detail certain neonatal movements, and Birdwhistell (e.g., 1) has presented a general system of movement analysis which may have relevance to the study of infants. These studies appear to be fair representatives of the research use of cinematographic techniques in the study of young infants. It may be that motion-picture photography is such an easy observational technique compared with alternative methods that the researcher collects thousands of feet of film with understandable gusto, only much later to face the question: What do I do with all these pictures? Descriptive and illustrative uses of film have been the most productive answers to date.

As part of an investigation of neonatal behavior, we set out to develop a technique which would permit the assignment of a numerical index to the movement of newborns. Furthermore, we wanted a technique which would have the advantage over stabilimetric devices on two counts: (a) in providing a fully comparable score, that is, a movement index that could be used in different research settings without introducing problems of mechanical comparability of instruments; and (b) in preserving a record of moment-to-moment changes in the behavior of the baby. For these purposes, motion-picture photography seemed appropriate; our problem was to devise a technique of film-analysis which would be reproducible, fully comparable across different settings, and productive of a meaningful measure.

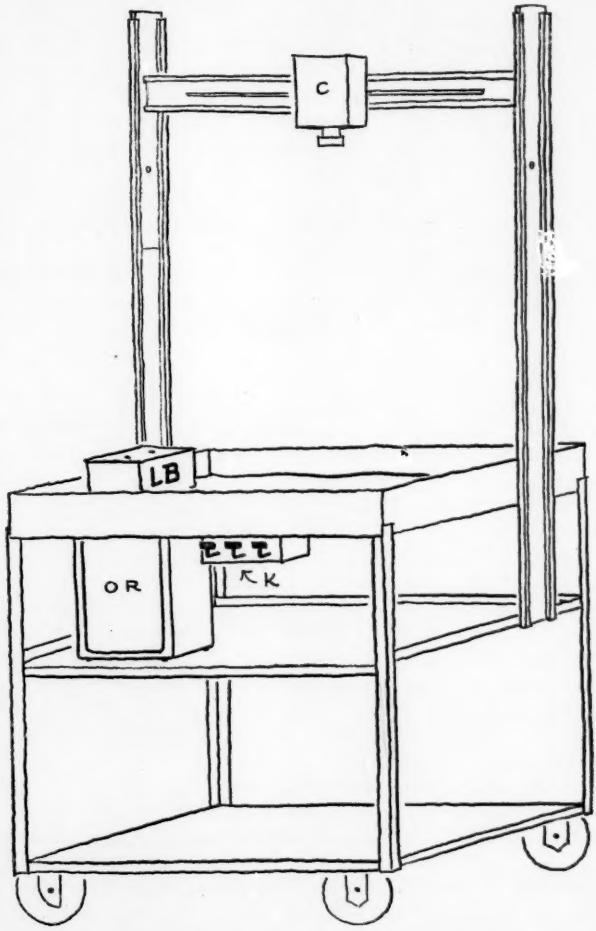

FIGURE 1—Observation cart in which infants were photographed. C: camera; K: telegraph keys; OR: Esterline-Angus operations recorder; LB: light-box used to synchronize OR and film.

Method

Setting and Subjects

Normal, full-term infants delivered in the maternity service of Grace-New Haven Community Hospital were observed during the lying-in period of five days. Fifteen of these newborns were photographed on each day of their hospital stay, 10 of them for five 30-sec. film strips daily, the remaining five for two 30-sec. strips daily. The babies were brought from the nursery to an unoccupied room on the same floor of the hospital, where they were photographed shortly before the midday feeding. Throughout the observation, the infants were dressed in hospital shirt and diapers. The period of observation, which included data-collection described in detail elsewhere (10), was from 20 to 30 min. in duration. The five 30-sec. filmstrips made for 10 babies were separated one from another by 90 sec.; the two filmstrips made for five babies were separated by 450 sec. Each strip was marked for later identification by photographing a series of letters and numbers coded to identify the child, the day of observation, and the strip.

Apparatus

The infants were placed supine on a mattress which fit an equipment cart adapted for photographic observation. Figure 1 is a schematic representation of this cart. An overhead girder approximately 3 ft. above the baby supported the motion-picture camera C (in this case, Bell and Howell 16-mm, Model DL-70). The use of a wide-angle lens (10-mm) provided a full view of the infant at relatively close placement of the camera. The camera was run at 16 frames per sec., and no artificial illumination other than the overhead lights already in place was used. A bank of telegraph keys (K) on each side of the cart permitted the recording of other data on an Esterline-Angus operations recorder (OR). Depression of certain keys also illuminated small indicator lights in a box (LB), which lights were used to check synchronization of film and recorder.

When the films had been processed, they were marked for analysis. The first frame of a filmstrip was located and marked by painting the preceding frame with ink. Then, a predetermined number of frames was counted off and another frame marked with a different color of ink. The most satisfactory method was to mark successive frames with ink of colors in the spectral order, i.e., red for the first frame, then orange, yellow, green, blue, and violet. The present technique was developed with a 5-sec. interval between frames selected for analysis (80 frames at 16/sec.), but other intervals can be used [*see* (15) for an analysis using a 1-sec. interval]. Thus, after the marking was completed, six frames from each 30-sec. filmstrip were marked for analysis—the first frame, the 81st (5 sec. from the start of the strip), the 161st (10 sec.), the 241st (15 sec.), the 321st (20 sec.) and the 401st (25 sec. from the beginning of the filmstrip).

Technique of Film-Analysis

The next step toward obtaining a reproducible measure of neonatal movement was to make a permanent copy of the filmframes selected for analysis. This step was taken by means of the filmanalyzer illustrated schematically in Figure 2. The original filmstrip (usually on a 100-ft. roll of film with several other strips) was placed on a reel (RI) and fed past a projector-lamp box (P) to another reel. This second reel was driven by a sewing machine motor (M) through a variable speed control device (R), so that a foot switch operated by the observer controlled the rate at which the film moved past the projector-lamp box. When a film frame selected for analysis came into position opposite the lamp, the film drive was stopped. The image of the infant was projected against a first-surface mirror (Mi) in a wooden chimney (C) and thence to a glass surface (G), 11 by 14 in.

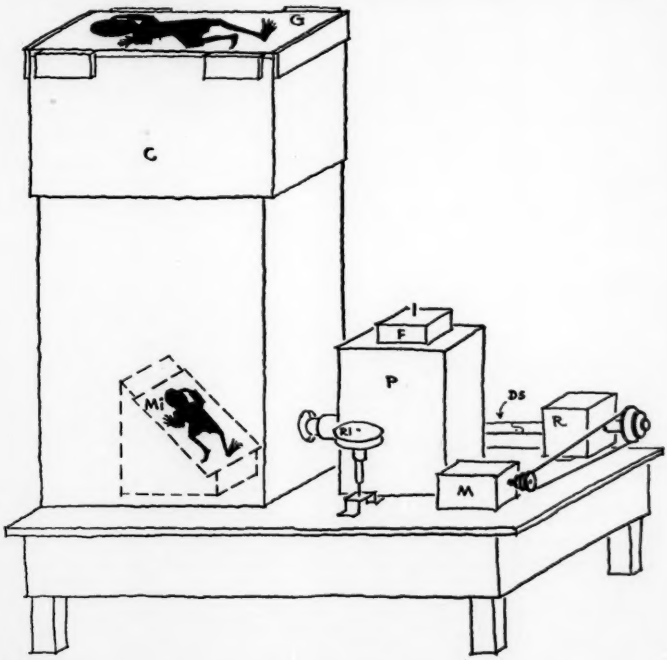

FIGURE 2—Schematic representation of filmanalyzer used in the present study. M: motor; R: variable-speed device; DS: driveshaft to takeup reel; RI: film reel; P: projector-lamp box; F: fan; Mi: first-surface mirror; C: enclosed wooden chimney; G: plate glass sheet.

in size. Draftsman's tracing paper was placed on this glass surface, and significant details of the image were traced for analysis. It should be noted once more that the particular technique for analysis and the particular dimensions of the apparatus used are not crucial in determining the measure of movement under study. Calibration procedures described below permit precise comparison of film analyses performed on instruments quite different from the one illustrated here.

Each of the filmframes selected for analysis was projected in the manner just described, and the position of the following parts of the infant's body was indicated by dots marked on tracing paper: visible ear, right palm, left palm, right knee, left knee, right ankle, left ankle. It was necessary also to mark on each sheet of tracing paper certain "landmarks," (usually the light box), so that one sheet could be accurately superimposed on another. Thus, for each filmframe selected for analysis (six per 30-sec. observation period, in this case), there was prepared a sheet of tracing paper which showed dots representing the position of seven body parts. Occasionally, the outline of the infant's body was sketched; occasionally, this outline was made in some detail. However, only the seven dots on each seet of tracing paper were used to obtain the movement measure.

Once the tracing-paper records of filmframes were made, the next task was to get a numerical index of changes in position of the body parts marked. A summary sheet was drawn, in which the dots representing limb positions from all relevant frames were shown, again coded in color to indicate which frame was involved. Then, the linear distance was measured from ear-mark (frame 1) to ear-mark (frame 81), from right-palm-mark (frame 81) to right-palm-mark (frame 161), and so on. Measurements of this order produced the protocol from which movement measure was extracted. An example of the protocol is shown in Table 1. As can be seen, there was now available an *index of displacement* for several parts of the infant's body, that is, a numerical representation of the shift in position of head, arms, and legs over a specified interval. It is important to note that this is not a measure of total movement, but rather a measure of displacement over a predetermined interval. The neonate may show wild flailing of arms and legs in the interval, only to return to exactly the same position by the time the next frame selected for analysis rolls by. We felt this possibility was unlikely enough to proceed with the development of the technique. Of course, shorter and shorter intervals could be used (e.g., examining every frame at high-speed photography) with an increase in approximation to a measure of total movement; correspondingly, the analytic task could be shortened by choosing a longer interval (e.g., a frame-sketch taken every 15 sec.). Our exploratory studies had led to a choice of the 5-sec. interval as optimally balanced between hopelessly difficult isomorphic representation of total movement and an impressionistic determination.

Thus, at the level of data-protocols, our index of displacement is reduced from "the real child" in three important ways: (a) it is based on film from

TABLE 1

REPRESENTATIVE EXAMPLE OF BASIC PROTOCOL

Subject A-1, *Day* 1, *Observation* 1 — *Analyzer:* ALS

	Frame-to-Frame Displacement of Image (cm.)				
	1–81	81–161	161–241	241–321	321–401
RH	.75	.65	.85	1.65	.90
LH	.30	6.15	.95	5.70	1.05
RK	3.05	1.70	.70	2.00	1.40
LK	3.40	1.25	4.30	3.10	1.95
RA	2.30	.60	.75	2.20	.60
LA	2.20	1.15	3.15	1.35	.40
(L)E	.10	.80	.30	.10	.75

NOTE.—R indicates *right*; L, *left*; H, *hand*; K, *knee*; A, *ankle*; E, *ear*.

a single fixed-position camera; (b) it involves only seven points on the body; (c) it is an index of displacement of parts, not of total movement.

Even the reduced protocol index is not easily manageable for all purposes, however, and for ease of computation, two further reductions were made. Displacements over the entire 30-sec. observation period were averaged to obtain a mean index of displacement, and a single index of movement per observation was obtained by adding the two palm and the two ankle measuses. Therefore, in most of the findings reported below, a single number per child per observation period is used, and it is this numeral which will be called the Movement Index (MI). More detailed examinations of displacement can be made, however, from the basic protocol illustrated in Table 1.

Calibration

In order to ascertain the true distance of displacement, a calibration was made of the camera-rig and filmanalyzer. A black paper strip, 50 cm. in length, was photographed in the position normally occupied by an infant on the observation cart, and the processed film of this strip was projected through the film analyzer and measured. The obtained length of the projected image of the strip was 18.30 cm., so that the protocol data need only to be multiplied by 50/18.30 or 2.73 to obtain a measure of true horizontal displacement of limbs over a 5-sec. interval. The resulting metric, an index of displacement in centimeters per limb over a 5-sec. period (cm./limb/5-sec.), although of no particular interest in itself, has the inestimable value of permitting numerical comparison of results from one research setting to another.

Reliability

The entire analytic procedure, from the marking of film to the presentation of a summary protocol, was performed on a randomly selected strip by two observers. The interobserver coefficient of agreement (r) was .96.

Application of the Method

The method of filmanalysis described in the foregoing pages has been and is being used in several studies of neonatal behavior, where its usefulness is being put to the central test, i.e., whether meaningful statements can be developed using it. Some of the results obtained will be reviewed here, together with some previously unreported findings, in order to demonstrate the flexibility and appropriateness of the MI measure.

Individual Differences

Stable individual differences over the five days of the lying-in period have been demonstrated for the movement measure, even when only one 30-sec. filmstrip per day was used in the analysis (10).

Day-to-Day Variation

The MI measure showed systematic and significant increases over the five-day lying-in period, confirming earlier observations of an increase in activity over the first days of life (10).

Laterality Variation

One of the advantages of a cinematographic technique is that it permits the accurate assessment of variations in laterality, a problem that has been of interest to psychologists of infancy for some time (e.g., 12). Figure 3 shows the values of MI for right and left palm, for right and left ankle, and for these values summed, over the lying-in period. As can be seen, there is no systematic superiority of one side over another during these first days of life. Nor does this representation of 15 infants blur underlying stable individual differences in sidedness. For each of the five days of observation, two inequalities can be determined—right palm displaced more or less than left palm, and right ankle displaced more or less than left ankle. Thus, for the entire period of study, each child will show a "sidedness" score from 0 (no instances of right palm or ankle displaced more than left palm or ankle) to 10 (right member displaced more than left in every instance). The obtained array of sidedness scores can be evaluated against a hypothesis of random variation of sidedness by matching the obtained results against expected values drawn from a binomial expansion. Table 2 shows the result of this comparison. It seems reasonable to conclude from this observation that newborn infants show no systematic variation in laterality of movement during the first five days of life.

The null hypothesis with respect to laterality will not hold if head position is examined. The MI tracings were scored for position of the head,

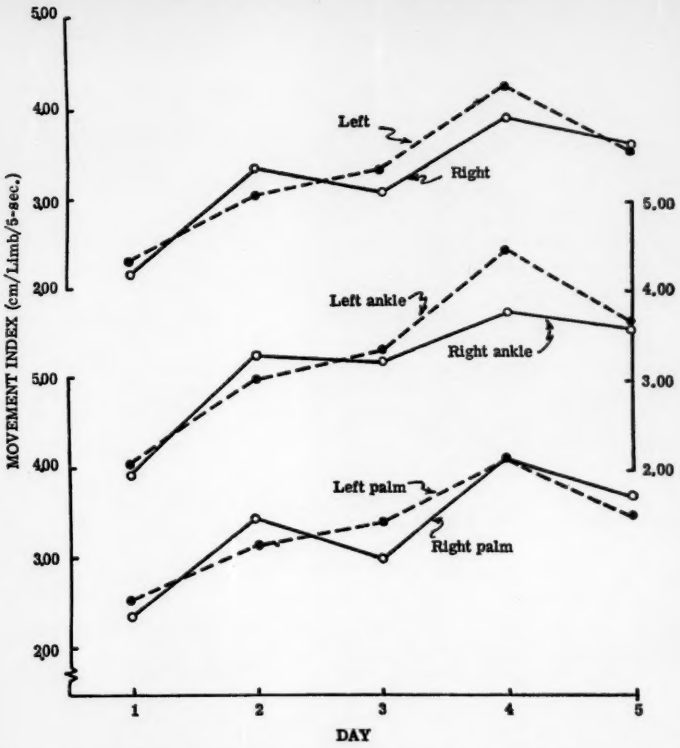

FIGURE 3—Displacement of limbs over the first five days of life. Lower curves show movement of palms; middle curves show movement of ankles; upper curves show the summation of the lower pairs of curves.

TABLE 2

NUMBER OF INFANTS SHOWING THREE DEGREES OF LATERALITY (OBTAINED) AND NUMBER TO BE EXPECTED ON A HYPOTHESIS OF RANDOM DISTRIBUTION OF PREFERENCE FOR RIGHT OR LEFT

	SIDEDNESS INDEX		
	0 1 2 3	4 5 6	7 8 9 10
Expected	2.6	9.8	2.6
Obtained	4	9	2

and a head-laterality determination made for each day of the hospital stay. In nine of 15 cases, the infant's head was to the right side (i.e., right cheek on or near the pad) *on every day,* although one of these children occasionally showed midline or left orientation. Of the remaining six infants, four were predominantly to the left only on one day, one was predominantly to the left on two days, and only one child had his head turned predominantly to the left on three of the five days in the hospital. We may conclude from these observations that the babies in our sample showed a strong tendency to keep their heads to the right.

Of more interest to the general issue of laterality variation, however, is the comparison of the laterality of MI in infants who show a strong head orientation toward the right with that of infants who show not-so-strong an orientation toward the right. Contrary to any simple hypothesis about laterality in neonates, there is no significant difference (either based on statistical considerations or based on visual inspection of the records) in the tendency to move right rather than left limbs between the two groups determined by head orientation. It is apparently not the case that, as the head leads, so the limbs will follow.

Summary

A method of observing newborn movement and a technical procedure for the analysis of motion-picture records have been described. The application of the technique has produced interesting and suggestive findings about individual variation in movement, about day-to-day changes in movement, and about laterality. These findings support the use of the method in the further exploration of the behavior in newborns, particularly in the study of response to systematic changes in the external environment of the child.

References

1. Birdwhistell, R. L. Background to kinesics. *Etc. Rev. gen. Semant.,* 1955, 13, 10-18.
2. Campbell, B. A. Design and reliability of a new activity-recording device. *J. comp. physiol. Psychol.,* 1954, 47, 90-92.
3. Escalona, S. K., & Leitch, M. Early phases of personality development: a non-normative study of infant behavior. *Monogr. Soc. Res. Child Develpm.,* 1952, 17, No. 1 (Serial No. 54).
4. Fries, M. E. Psychosomatic relationships between mother and infant. *Psychosom. Med.,* 1944, 6, 159-162.
5. Irwin, O. C. The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monogr.,* 1930, 8, 1-92.

6. Irwin, O. C. The distribution of the amount of motility in young infants between two nursing periods. *J. comp. Psychol.*, 1932, 14, 429-445.

7. Gesell, A. *An atlas of infant behavior.* Yale Univer. Press, 1934.

8. Gesell, A., & Halverson, H. M. The daily maturation of infant behavior. A cinema study of postures, movements and laterality. *J. genet. Psychol.*, 1942, 61, 3-32.

9. Gilmer, B. v.H. An analysis of the spontaneous responses of the newborn infant. *J. genet. Psychol.*, 1933, 42, 392-405.

10. Kessen, W., Williams, E. J., & Williams, J. P. Selection and test of response measures in the study of the human newborn. *Child Develpm.*, 1961, 32, 7-24.

11. Marquis, D. P. Can conditioned responses be established in the newborn infant? *J. genet. Psychol.*, 1931, 39, 479-492.

12. Marquis, D. P. A study of activity and postures in infants' sleep. *J. genet. Psychol.*, 1933, 42, 51-69.

13. Murphy, L. B. *Personality in young children.* Vol. 1. Basic Books, 1956.

14. Pratt, K. C., Nelson, A. K., & Sun, K. H. The behavior of the newborn infant. *Ohio State Univer. Stud., Cont. Psychol.*, 1930, No. 10.

15. Williams, J. P., & Kessen, W. Effect of hand-mouth contacting on neonatal movement. *Child Develpm.*, in press.

16. Wolf, K. M. The origins of individuality. Mimeographed book, 1957.

17. Wright, H. F. Observational child study. In P. H. Mussen, (Ed.), *Handbook of research methods in child development.* Wiley, 1960. Pp. 71-139.

AMPLITUDE OF AN INSTRUMENTAL RESPONSE FOLLOWING THE CESSATION OF REWARD[1]

RUTH B. HOLTON*
State University of Iowa

Several investigators (1, 4, 6, 7, 8) have reported increments in some measures of response strength following the omission of a customary reward. Few studies, however, have demonstrated variations in the magnitude of these increments as a function of experimental variables.

Roussel (7) reports data which suggest that amount of previous training is positively related to increased running speed in rats following nonreward. Haner and Brown (3) found children exerted more force when blocking was introduced near a goal than when it occurred at points farther removed from the goal. These findings suggest that increments in the amplitude of an instrumental response, occurring as a consequence of nonreinforcement, may be affected by level of previous reinforced training and distance of the point of nonreward from the goal. The study reported here was designed to test these hypotheses with young children.

A marble reinforcement was delivered each time *S* responded to (pushed) the correct one of two stimulus windows. The number of times *S* pushed the

---

* Deceased.

[1] This article is based upon a Ph.D. dissertation submitted to the faculty of the Iowa Child Welfare Research Station, State University of Iowa. The author is indebted to Dr. Charles C. Spiker for directing the research and to Mr. John Peterson for constructing the apparatus.

Child Develpm., 1961, 32, 107-116.

correct stimulus window constituted level of reinforced training. Distance from the goal was expressed in terms of the number of marbles necessary for *S* to finish filling a marble board, the completion of which entitled him to a toy reward. Marble reinforcements were withheld after two levels of reinforced training and at two distances from the goal.

It was expected that, with goal distance held constant, *Ss* having a relatively high number of reinforced trials prior to the failure of reward would show greater increments in response amplitude than *Ss* having fewer reinforcements. Furthermore, with level of reinforcement constant, it was expected that *Ss* nonrewarded close to the goal would show greater increments in response amplitude than *Ss* nonrewarded far from the goal.

Method

Subjects

Forty-five children, ranging in age from 3 years, 8 months, to 5 years, 6 months, drawn from the Iowa Child Welfare Research Station Laboratories served as *Ss*. *Ss* were assigned at random to three groups of 15 each, with the stipulation that each group have the same proportion of boys to girls.

Apparatus

The apparatus was mounted in a window, 33 in. from the floor, which separated an observation booth from an experimental room. Only the face of the apparatus was visible to *S*, who responded from the experimental room, while *E* recorded responses and operated the necessary controls from the booth.

The face of the apparatus consisted of a black wooden panel, 24 by 16 in. Extending 2½ in. out from the panel were two plexiglass windows 5 in. square, behind which were placed two identical orange stimulus cards. Mounted between the two windows, which were 6¾ in. apart, was a vertically placed 11-in. transparent plastic tube filled with marbles and a marble dispenser which fed into a receiving cup. A rod extending across the front of the apparatus held a curtain which could be drawn to conceal the apparatus from *S*'s view.

Each window was mounted on the weighing platform of a Hanson Utility Scale (Model No. 3000—25 lb. capacity) in which the recording mechanism had been disengaged from the return spring, allowing the scale pointer to remain at the point of maximum push until returned to the zero position by *E*. A microswitch, attached to the spring of each scale, activated a marble release mechanism when a force of ½ lb. or more was delivered to the correct window. A single push against either window cut off the circuit until it was manually reset by *E*; thus, *S* was prevented from obtaining more than one marble per trial or from obtaining a marble by pushing the correct window after pushing the incorrect one.

Additional materials included a wooden stand on which *S* stood during the experiment and a 32-in. marble board with 11 holes. A variety of dime-store toys served as rewards for filling the marble board.

Procedure

S was brought into the experimental room and the apparatus was introduced as a game in which he would be given an opportunity to win a prize. He was told that each time the curtain opened he was to push one of the windows and that, if he pushed the correct one, a marble would drop into the cup, but if he pushed the incorrect one, the marble would not be delivered and he would have to wait until the curtain opened the next time before pushing again.

E then left the room and started pretraining by opening the curtain from the control booth. As soon as *S* pushed one of the windows, the curtain was closed and the window to which the response had been made and the amplitude of that response were recorded. The reopening of the curtain served as a signal for the beginning of the next trial. *Ss* in one group were pretrained until they had made 4 correct responses; a second group, 17 correct responses; and the third group, 22 correct responses.[2]

Following pretraining, *E* returned to the experimental room and placed a board with 11 holes on the ledge below the apparatus, directly in front of *S*. *S* was told that the way to win a prize was to earn enough additional marbles to fill all the holes in the marble board. Before training was resumed, *S* selected the toy he wished to win from a display of dime-store toys.

When *S* had filled a predetermined number of holes in the board, marble reinforcements were withheld for responses to either window for a period of seven trials. Following these seven trials, responses to the window which had been incorrect prior to the failure of reward were reinforced until *S* had filled the remaining holes in the board and had won his prize.

One treatment group, 13-N, was introduced to nonreward near the end of the marble board after 13 reinforced training trials. A second group, 26-N, also experienced nonreward near the end of the board but after 26 reinforced training trials. The third group, 26-F, received the larger number of reinforced training trials and was nonrewarded far from the goal. Table 1 summarizes the conditions under which nonreinforcement was introduced for the three groups. It will be noted that group 26-N received twice as many reinforcements prior to nonreward as group 13-N, with distance from the goal held constant. Groups 26-N and 26-F, on the other hand, received an equal number of reinforcements but experienced nonreinforcement at different distances from the goal.

Within each of the three groups, eight *Ss* were reinforced for responding to the left window and seven *Ss* for responding to the right window. In addition to the 45 *Ss* employed, one *S* was omitted from group 13-N due

[2] Varying the number of correct responses made during pretraining made it possible to employ the same 11-hole marble board for all three groups.

TABLE 1

CONDITIONS UNDER WHICH NONREWARD WAS INTRODUCED FOR THE THREE TREATMENT GROUPS

	Level of Reinforced Training			
Group	*Correct Responses During Pretraining*	*Correct Responses During Marble Board Training*	*Total Number of Correct Responses*	*Marbles Needed to Reach Goal*
13-N	4	9	13	3
26-N	17	9	26	2
26-F	22	4	26	7

to an experimenter error in procedure. One *S* in group 26-N and two in group 26-F were eliminated because of apparatus failure. Three more *S*s, one in group 26-N, and two in group 26-F, were eliminated during pretraining when they persisted in responding either alternately or randomly to the incorrect window.

RESULTS

*S*s made few errors in learning to respond to the correct stimulus window. Groups 26-N and 26-F, in reaching a criterion of 26 correct responses, made 2.46 and 3.60 mean errors, respectively. These means did not differ significantly. Group 13-N, in making 13 correct responses, made .80 mean errors. All errors occurred early in pretraining; thus, all *S*s were responding consistently to the reinforced window prior to the omission of reward.

In order to provide a stable estimate of changes in amplitude occurring as a consequence of nonreinforcement, two scores were obtained for each *S*, the mean force exerted on the four trials preceding the trial on which nonreward was introduced (prescore), and the mean force exerted on the four trials following the first nonreward trial (postscore).[3] Mean pre- and postscores for each group are graphically presented in Figure 1. Inasmuch as there was a tendency for the distributions of these scores to be positively skewed with positively correlated means and variances, the statistical analyses reported below were performed upon a square root transformation of these scores.[4] Means and SDs of the transformed scores are presented in Table 2.

[3] Amplitudes recorded for the first nonreinforced trial were not included in the postscore since some *S*s seemed aware that the marble was not being delivered before they had completed the push (making this, effectively, a nonreward trial), while other *S*s did not appear aware that the marble was being withheld until the push had been completed (making this a reward trial).

[4] Employing the 5 per cent level as the coefficient of risk, *F* tests carried out on the nontransformed data led to the same conclusions regarding the experimental hypotheses as those given below.

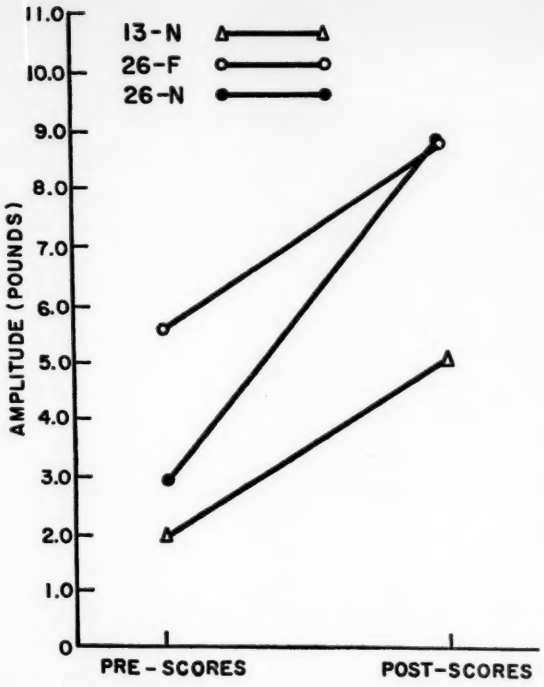

FIGURE 1—Mean force exerted on the four trials prior to the first nonreward trial and mean force exerted on the four trials following the first nonreward trial.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF TRANSFORMED PRE- AND POSTSCORES FOR THE THREE TREATMENT GROUPS

	PRESCORES		POSTSCORES	
Group	*Mean*	*SD*	*Mean*	*SD*
13-N	1.38	.28	2.22	.34
26-N	1.65	.43	2.93	.54
26-F	2.23	.78	2.85	.84

Performance Prior to Nonreward

An analysis was first made of the transformed prescores. A factorial analysis was conducted, in which the main effects were treatment and position of correct window. Only the treatment effect was significant ($p < .001$), indicating that there were reliable differences in the force exerted by the three groups prior to the introduction of nonreward. Separate analyses, comparing groups 26-N and 13-N and groups 26-N and 26-F, indicated that the force exerted by groups 26-N and 13-N did not differ significantly, but that group 26-F was pushing significantly harder than group 26-N prior to nonreward.

Change in Performance as a Consequence of Nonreward

It is apparent in Figure 1 that each group showed a substantial increase in amplitude following the introduction of nonreward. To investigate, statistically, the characteristics of these increases, an analysis was made of the transformed pre- and postscores. A Lindquist Type III analysis (5) was conducted, in which the main effects were treatment, position of correct window, and pre- and postscores.[5] The treatment and the pre-post main effects and the pre-post by treatment interaction were all highly significant, indicating that the groups differed in mean force exerted, that there was an increase in amplitude following the introduction of nonreward, and that the magnitude of this increase varied with experimental treatment.

To test the specific hypotheses that increases in amplitude would be greater for group 26-N than for group 13-N and greater for group 26-N than for group 26-F, separate Type III analyses were conducted, employing the error terms obtained in the over-all analysis. The analysis comparing groups 26-N and 13-N is summarized in Table 3. The results indicate that more force was exerted by group 26-N with both groups showing a significant increase in force following the omission of reward. Of primary interest, however, is the pre-post by treatment interaction which indicates that group 26-N showed a significantly greater increment in amplitude than group 13-N. The significant interaction between pre-post scores, treatment, and position reflects greater differential increases in force, as a function of treatment, for *S*s responding to the right stimulus window.

The analysis comparing groups 26-N and 26-F is also summarized in Table 3. The results show a significant increase in amplitude following the omission of reward for both groups, with a significantly greater increase occurring for group 26-N.

An examination of the pre- and postscores for all 45 *S*s showed that 44 of them showed an increase in response amplitude following the omission of reward. The 45th *S* showed no change.

---

[5] Since the major hypotheses of this experiment concern the pre-post by treatment interactions, and since the proper test of this effect assumes homogeneity of the pre-post by subject interactions, Bartlett's test for homogeneity was applied to the variances of the differences between the transformed pre- and postscores for each treatment group. The resultant χ^2 was 3.86 ($p > .10$).

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF TRANSFORMED PRE- AND POST-SCORES FOR GROUPS 26-N AND 13-N AND FOR GROUPS 26-N AND 26-F

		GROUP 26-N vs. GROUP 13-N			GROUP 26-N vs. GROUP 26-F		
Source	*df*	*MS*	*F*	*p*	*MS*	*F*	*p*
Subjects	29						
Treatment (T) ..	1	3.67	6.12	<.025	.92	1.53	>.20
Position (P)	1	.35			1.97	3.28	>.10
T × P	1	.76	1.27	>.20	.01		
error (b)*	36	.60			.60		
Within Subjects ...	30						
Pre-Post (PP) ...	1	16.87	153.36	<.001	13.71	124.64	<.001
PP × T	1	.75	6.82	<.025	1.61	14.64	<.001
PP × P	1	0.00			.11		
PP × P × T ...	1	.51	4.64	<.05	.13	1.18	>.20
error (w)*	36	.11			.11		

* Error terms are those obtained in over-all analysis.

Performance as a Function of Successive Nonreward Trials

Although the largest increments in amplitude occurred on the trial following the introduction of nonreward, amplitude measures for all seven nonreward trials suggest that amplitude continued to increase with successive nonreinforcements. The statistical significance of this trend was tested by a Lindquist Type I analysis (5). A highly significant trial effect ($p <$ .001) indicated that force did continue to increase with successive nonreward trials.

Persistence of the Originally Correct Response

Since *S*s continued to respond until they had won enough marbles by pushing the previously incorrect window, it was possible to obtain a measure of persistence of the response to the originally correct window. The number of responses made following the introduction of nonreward, before *S*s responded to the window which had been incorrect during training, constituted the persistence measure. The mean persistence score for group 26-N was 7.20, for group 26-F, 4.00, and for group 13-N, 6.20. The distributions of these scores were extremely variable, however, and differences among the groups were not statistically significant.

Discussion

The results indicate that *Ss* given a relatively large number of reinforcements and nonrewarded near the goal showed a greater increase in amplitude than (a) *Ss* given fewer reinforcements and nonrewarded near the goal and (b) *Ss* given a large number of reinforcements and nonrewarded far from the goal. It should be recalled, however, that group 26-F was pushing harder prior to nonreward than group 26-N. This suggests the possibility that the differential increase in amplitude occurring for the two groups given the same number of reinforcements but nonrewarded at different distances from the goal was a consequence of a ceiling effect imposed upon the performance of group 26-F. That is, some *Ss* in group 26-F may have been responding near maximum prior to nonreward and were thus physically limited in the amount of increment that could accrue as a consequence of nonreinforcement.

The data offer two lines of evidence relevant to such an interpretation. First, if a ceiling effect did act to limit the increment in amplitude occurring for group 26-F, one might expect this group to show a negative correlation between prescores and amount of increase in amplitude following nonreward. The correlation between transformed prescores and amount of increase (obtained by subtracting transformed prescores from transformed postscores) was $-.01$. There was no significant tendency, therefore, for high amplitude scores prior to nonreward to be associated with small increases in amplitude following nonreward. Secondly, it was possible to match seven *Ss* in group 26-N within ½ lb. of the prescores of seven *Ss* in group 26-F. Amount of increase in amplitude was compared for these *Ss*. A t value of 4.56 ($6\ df$, $p < .01$) indicated a significantly greater increase in amplitude for *Ss* in group 26-N than for *Ss* in group 26-F. This evidence appears to justify the conclusion that the differential change in amplitude occurring for groups 26-N and 26-F was a consequence of experimental manipulation of goal proximity, rather than a consequence of differences in amplitude level prior to the introduction of nonreward.

The tendency for amplitude to increase with successive nonreward trials suggests that the effect of nonreward upon performance was, up to some point, cumulative. It is of interest to note in this connection that *Ss* who persisted for some trials in responding to the originally correct window (the nonreinforced one after the failure of reward) tended to maintain a high level of amplitude, if not increase it, for several trials following the seven-trial nonreinforcement period.

The fact that 44 of 45 *Ss* showed an increase in amplitude following the omission of reward accords with the hypothesis (2) that omission of a customary reward gives rise to an emotional state (frustration) which increases drive and hence level of performance. The finding that such increases in performance are positively related to number of reinforcements received prior to nonreward and the proximity of the point of nonreward from the

goal offers support for the suggestion (9) that degree of frustration is some function of the amount of "reward expectation" which has been developed.

Summary

Three groups of preschool children learned to push the correct one of two stimulus windows for a marble reinforcement. After a predetermined number of correct responses had been made, reinforcement was omitted for seven trials. After these nonreward trials, responses to the window which had been incorrect during training were rewarded until *S* had obtained enough marbles to finish filling a marble board, entitling him to a toy reward.

Mean force exerted on the four trials prior to the introduction of nonreward was compared with the mean force exerted on the four trials following the introduction of nonreward in order to determine the effect of reward failure upon increases in amplitude of response.

Results showed that, with goal distance controlled, a group receiving a relatively high number of reinforcements prior to nonreward showed a significantly greater increase in amplitude following nonreinforcement than a group receiving fewer reinforcements. Furthermore, with number of reinforcements controlled, a group introduced to nonreward close to the completion of the marble board showed a significantly greater increase in amplitude than a group experiencing nonreward at a point farther removed from the goal. Forty-four of the total 45 *S*s showed an increase in amplitude following the omission of reinforcement.

It was suggested that the experimental variables subjected to manipulation in this investigation may have differentially influenced the degree of frustration resulting from the omission of a customary reward.

References

1. Amsel, A., & Roussel, J. Motivation properties of frustration: effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, 43, 363-368.
2. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, 48, 465-495.
3. Haner, C. F., & Brown, P. A. Clarification of the instigation to action concept in the frustration-aggression hypothesis. *J. abnorm. soc. Psychol.*, 1955, 51, 204-206.
4. Hilgard, E., & Marquis, D. *Conditioning and learning*. Appleton-Century-Crofts, 1940.
5. Lindquist, E. *Design and analysis of experiments in psychology and education*. Houghton Mifflin, 1953.

6. Marzocco, F. N. Frustration effect as a function of drive level, habit strength, and distribution of trials during extinction. Unpublished doctoral dissertation, State Univer. of Iowa, 1951.

7. Roussel, J. Frustration effect as a function of repeated nonreinforcements and as a function of the consistency of reinforcement prior to the introduction of nonreinforcement. Unpublished master's thesis, Tulane Univer., 1952.

8. Skinner, B. F. *Behavior of organisms.* Appleton-Century-Crofts, 1938.

9. Wilson, W., Weiss, E. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.,* 1955, 50, 51-60.

RELATIVE EFFECTIVENESS OF DIFFERENT REINFORCEMENT COMBINATIONS ON CONCEPT LEARNING OF CHILDREN AT TWO DEVELOPMENTAL LEVELS

WILLIAM J. MEYER* *and* STANLEY B. SEIDMAN
University of Pittsburgh

In a recent paper (7) the writers attempted to demonstrate that under certain conditions developmental level is an important variable in the determination of the effectiveness of various types of reinforcements. Reasoning that under relatively ambiguous learning conditions older *Ss* would be sensitive to a larger variety of reinforcing cues, an experiment was designed in which the *Ss* received no prior instructions concerning the nature or meaning of the reinforcements employed. The reinforcement combinations were as follows: *Right-Wrong* (RW), *Right-Nothing* (RN), *Nothing-Wrong* (NW), and *Buzzer-Nothing* (BN). It was assumed that the presence of the ***nothing*** condition in three of the combinations would provide sufficient ambiguity to permit differential interpretation of the examiner's behavior and that such differences might be related to developmental level. It was further anticipated that the presentation of the buzzer without specifying its meaning, i.e., the buzzer could either signify a correct or an incorrect response, would also provide an ambiguous reinforcement. Since little is known about the nature of extraneous reinforcement cues, no formal predictions concerning the experiment were attempted.

Experimental verification was attempted by randomly placing 20 youngsters in each of the four treatment groups and at each of two developmental levels, four- and five-year-olds and eight- and nine-year-olds, affording a total of 160 *Ss*. The task consisted of a simple discrimination problem, learning to respond to the larger of two blocks. An analysis of variance of the acquisition data revealed a significant trials $\times$ age interaction and a significant trials $\times$ age $\times$ treatments interaction. On the basis of these results it was concluded that the effectiveness of the four reinforcement combinations was related to age and that rate of acquisition was a function of age and treatments. Multiple comparison tests failed to yield significant differences, but inspection of the learning curves suggested that the NW combination was more effective for the older as compared with the younger *Ss* whereas the RW and BN combinations were more effective for the younger than the older *Ss*. Since the multiple comparison tests were not

* Department of Psychology, University of Pittsburgh, Pittsburgh 13.

Child Develpm., 1961, 32, 117-127.

statistically significant, it was decided to replicate part of the original study by comparing the RW-NW combinations in one experiment and the NW-BN combinations in a second experiment.

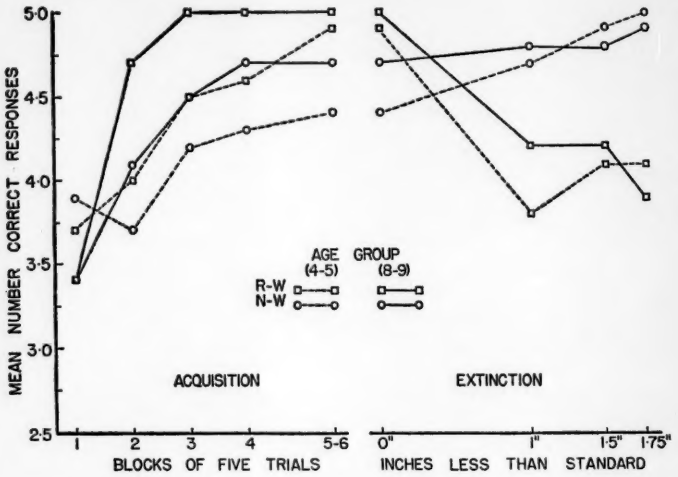

FIGURE 1—Mean number of correct responses on acquisition and extinction series for RW and NW treatments.

EXPERIMENT I

Procedure

The first experiment compared the RW-NW combinations. A total of 40 *Ss*, 20 four- and five-year-olds and 20 eight- and nine-year-olds attending a suburban public school, constituted the sample.[1] The *Ss* were from an average socioeconomic environment and were representative of an average intellectual population.

Ten *Ss* were randomly assigned to each of the four experimental groups comprised of two treatments and two developmental levels. The *Ss* in the RW group were told "right" following a correct response and "wrong" following an incorrect response. The NW group was told *nothing* following a correct response and "wrong" following an incorrect response.

The stimuli for the learning problem were identical to those employed in the earlier study. They comprised nine pairs of three-dimensional geometric blocks in the shape of cylinders, triangles, and rectangles. The color (red) and heights (2 in.) of all blocks were identical; only the basal surface

[1] The writers wish to express their appreciation to Mr. Allen Harris, supervising principal of the Eastmont School, for his cooperation in making *Ss* available for this study.

area varied from 1 to 2 in. These stimuli comprised the acquisition series. For the extinction trials, the same basal area relationships (2 : 1) were maintained, but the heights of the stimuli were reduced by 1 in., 1.5 in., and 1.75 in. These stimuli constituted the extinction series. The stimuli were presented in pairs; the position of the correct choice was randomized with the larger basal area always being the correct block. Additional apparatus consisted of a portable box that had a board attached to it which served as a blind.

The *Ss* were tested individually. Each *S* received the same instructions: "I am going to show you two blocks at a time. Each time you must choose, by pointing with your finger, the block that you think is the correct one. Remember that one block is always the right one to choose and one block is always the wrong one to choose." After reaching a criterion of 9 out of 10 correct responses or a maximum of 30 trials, whichever came first, each *S* was immediately started on the extinction series of five trials for each of the stimulus conditions. During this series no reinforcements were given.

Results

In the previous study it was reported that little learning occurred after the twentieth trial, hence, only the first 20 trials were included in the statistical analysis. A similar result was found in the present study and was handled in the same way. The data were organized into four blocks of five trials and are presented graphically in Figure 1. Table 1 shows the mean number of correct responses made by the two experimental groups on the 20 trials.

TABLE 1

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON ACQUISITION SERIES

	Age 4 to 5		Age 8 to 9	
Treatment	*Mean*	*Sigma*	*Mean*	*Sigma*
NW	16.10	3.06	16.70	3.26
RW	16.80	2.92	18.10	4.24

A summary of the analysis of variance for repeated measures on the same *Ss*, as described by Edwards (6), for the acquisition data is presented in Table 2. The Bartlett test was statistically nonsignificant. The following conclusions are warranted: (a) effects attributable to age, treatments, and the treatments × age interaction are not statistically significant; (b) the between trials effect is statistically significant, i.e., the learning curves are not flat; (c) the trials × treatments, trials × age, and the trials × age × treatments interactions fail to reach an acceptable level of significance. The

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF ACQUISITION DATA FOR RW-NW TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	2.25	1.87
Age (A)	1	2.75	2.29
T × A	1	.31	
Error A	36	1.20	
Trials	3	9.15	12.20**
Trials × T	3	1.86	2.48
Trials × A	3	.63	
Trials × T × A	3	.03	
Error B	108	.75	

** $p < .01$.

results of the foregoing analysis permit the conclusion that both the RW and NW reinforcement combinations are equally effective at both levels of development.

The data on mean performance of each experimental group on the three sets of stimuli constituting the extinction series are presented in Table 3 and the right side of Figure 1. The analysis of variance relevant to these data is presented in Table 4. This analysis indicates that over all the stimulus pairs there is a statistically significant difference in which the NW treatment is superior to the RW. The significant stimuli effect can be interpreted as meaning that the number of correct responses on each set of stimuli are significantly different. The significant stimuli × treatments interaction indicates that the slopes of the extinction curves are different. Inspection of Figure 1 shows that the NW group tends to improve in performance somewhat whereas the RW group declines.

TABLE 3

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON EXTINCTION SERIES

Stimulus Height (in. less than standard)	TREATMENT N-W				TREATMENT R-W			
	AGE 4 TO 5		AGE 8 TO 9		AGE 4 TO 5		AGE 8 TO 9	
	Mean	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*
0	4.40	.96	4.70	.94	4.90	.32	5.00	0.00
1	4.70	.67	4.80	.63	3.80	1.13	4.20	1.13
1.50	4.90	.32	4.80	.63	4.10	.87	4.20	1.61
1.75	5.00	.00	4.90	.32	4.10	.87	3.90	1.65

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF EXTINCTION-GENERALIZATION DATA FOR RW-NW TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	10.00	5.13*
Ages (A)	1	.12	
T × A	1	.03	
Error A	36	1.95	
Stimuli (S)	3	4.20	10.00**
S × T	3	3.72	8.85**
S × A	3	.34	
S × T × A	3	.14	
Error B	108	.42	

* $p < .05$.
** $p < .01$.

On the basis of our previous research it was expected that the NW treatment would be more effective with the older *Ss* whereas the RW treatment would be more effective with the younger *Ss*. The statistically nonsignificant treatments × age interaction clearly shows that these pairs of reinforcers are equally effective with both age groups. Though the nonsignificant treatments effect precludes any definitive statements concerning the reinforcement combinations, it should be noted that the RW condition produced the most rapid learning amongst the older *Ss* whereas for the same age group in the earlier study the NW treatment produced the most rapid learning. We shall offer an hypothesis later in this paper in an effort to explain this discrepancy.

In terms of the extinction data, the earlier study suggested that the NW group would show little if any decrement in performance whereas the RW *Ss* would demonstrate the typical extinction curve. The present data clearly support this expectation.

The results of the acquisition phase of this experiment are in agreement with the findings reported by Buss *et al.* (4) and Buss and Buss (5). These investigators, employing student nurses and psychiatric patients as *Ss*, found little difference in learning between their NW and RW groups. They interpreted their results as indicating that *right* is a much weaker reinforcer than *wrong* and that the critical component of the combinations is *wrong*. Our results are consistent with this notion and indicate that the differential strength of *right* and *wrong* exists at least as early as the fifth year of life.

Inconsistent with the Buss study, but consistent with Buchwald (3), our NW group performed better during extinction. In our earlier paper, wherein a RN group was included, the foregoing data were interpreted as

meaning that *nothing* when paired with *wrong* acquires positive reinforcing value but when paired with *right* acquires negative reinforcing value. The RW treatment, which obviously does not give the *S*s any cues concerning *nothing,* presumably should not affect its reinforcement value. Both Buss (4) and Buchwald (2, 3) found no differences between NW and RW on extinction and so support the foregoing assertions. The findings of the present study are clearly not in accord with these investigators. In interpreting our data, we take the position that whenever *right* is explicit in a pair of reinforcers, a subsequent failure of this response to occur has a negatively reinforcing effect. A similar effect is not derived from the *wrong* condition because, except for the first 10 trials or so, the *S*s are essentially functioning as an RN group wherein *nothing* acquires negative reinforcement value.

Experiment II

Procedure

The second experiment compared the NW and BN combinations. A total of 40 *S*s, 20 four- and five-year-olds and 20 eight- and nine-year-olds attending a suburban public school, constituted the sample. The *S*s were from the same school as those in Experiment I.[2] The procedures and stimulus materials were identical to those in Experiment I with the exception that the RW group was replaced by the Buzzer-Nothing (BN) group. This latter group heard a buzzer (which was secured to the board and out of sight of the *S*s) every time they made a correct response and *nothing* for an incorrect response. No explanation of the meaning of the buzzer was given.

Table 5

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON ACQUISITION SERIES

	Age 4 to 5		Age 8 to 9	
Treatment	*Mean*	*Sigma*	*Mean*	*Sigma*
NW	16.10	3.06	16.70	3.26
BN	13.60	5.31	13.80	3.02

Results

The relevant data for Experiment II are reported in Table 5 and Figure 2. The data were treated as before, and again the Bartlett tests were not statistically significant. The analysis of variance presented in Table 6

[2] The NW *S*s for the second experiment are the same as those used in Experiment I.

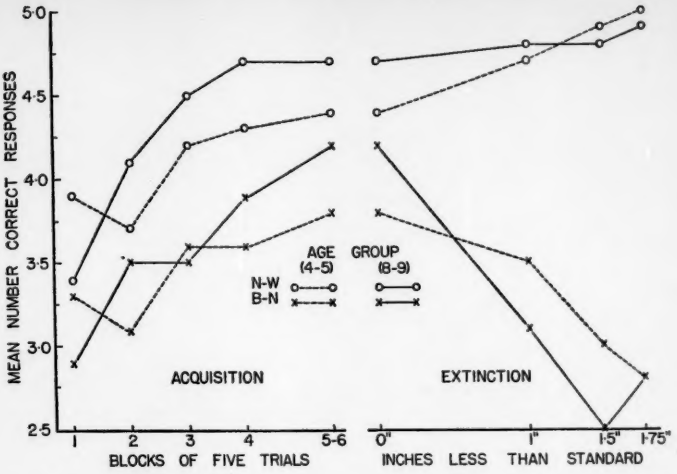

FIGURE 2—Mean number of correct responses on acquisition and extinction series for NW and BN treatments.

permits the following conclusions: (a) there is a significant difference between the two treatments in which the NW group is superior to the BN group; (b) both age groups perform similarly under the two treatments; (c) the nonsignificant treatments × age interaction indicates that the NW treatment is superior for both age groups.

TABLE 6

SUMMARY OF ANALYSIS OF VARIANCE OF ACQUISITION DATA FOR NW-BN TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	18.23	5.09*
Ages (A)	1	.40	
T × A	1	.10	
Error A	36	3.58	
Trials	3	4.58	7.05**
Trials × T	3	.14	
Trials × A	3	1.51	2.32
Trials × A × T	3	.12	
Error B	108	.65	

* $p < .05$.
** $p < .01$.

TABLE 7

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON EXTINCTION SERIES

Stimulus Height (in. less than standard)	TREATMENT N-W				TREATMENT B-N			
	AGE 4 TO 5		AGE 8 TO 9		AGE 4 TO 5		AGE 8 TO 9	
	Mean	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*
0	4.40	.96	4.70	.94	3.80	1.31	4.20	1.31
1	4.70	.67	4.80	.63	3.50	1.71	3.10	1.28
1.50	4.90	.32	4.80	.63	3.00	1.69	2.50	1.26
1.75	5.00	.00	4.90	.32	2.80	1.54	2.80	1.47

The data relevant to the extinction series for this experiment are presented in Table 7 and on the right side of Figure 2. As was anticipated, the performance of the NW groups is clearly superior. The highly significant treatments effect shown in Table 8 indicates that the NW group in contrast with the BN group gave reliably more correct responses. In addition, the significant stimuli $\times$ treatments interaction shows that the slopes of these curves are different with the BN *Ss* declining sharply in performance after the discontinuation of reinforcement.

The results of our first study suggested that the BN treatment would be more effective with the younger *Ss* and the NW treatment would be more effective with the older Ss. The nonsignificant age $\times$ treatments interaction indicates that each age group responds similarly to the two treatments and therefore fails to support the earlier results. In this study the BN *Ss* of both

TABLE 8

SUMMARY OF ANALYSIS OF VARIANCE OF EXTINCTION-GENERALIZATION DATA FOR NW-BN TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	97.66	27.66**
Ages (A)	1	.06	
T $\times$ A	1	.30	
Error A	36	3.53	
Stimuli (S)	3	1.75	3.44*
S $\times$ T	3	5.57	10.71**
S $\times$ A	3	.77	1.28
S $\times$ T $\times$ A	3	.26	
Error B	108	.52	

* $p < .05$.
** $p < .01$.

age groups performed in a manner similar to the Right-Nothing group included in the original paper and suggests that the buzzer (when undefined) is no more effective than saying "right." The extinction data clearly support the previous findings and would not seem to require further comment here.

Discussion

The failure to replicate the age × treatments interaction requires examination. Since the experimental procedures were identical in every detail throughout the entire series of studies, it does not seem likely that this area of examination would prove fruitful. One possible source of variance perhaps exists in the samples employed in both studies. The *S*s in the earlier study were drawn primarily from a private university elementary school in which the general intellectual level is high (mean IQ = 124) and wherein considerable emphasis is placed on academic achievement and competition. In contrast, the second sample was drawn entirely from a public school wherein the general intellectual level was lower and where the school placed less emphasis upon academic competition. One might speculate that differences in achievement motivation were responsible for the divergent findings. However, we believe that until further data can be brought to bear on this issue the most plausible explanation of the failure to replicate is a statistical consideration. It seems reasonable that the significant interaction resulted from some slight differences in treatments effectiveness which when pooled for eight groups (four treatments and two developmental levels) achieved a magnitude sufficient for statistical significance and resulted in a Type I error.

Summarizing the results of the present study and the earlier report, it can be concluded that for both preschool age children and primary grade children verbal reinforcement combinations which include *wrong* as a component are more effective than just positive reinforcement (RN and BN). At both developmental levels *wrong* appears to be a stronger negative reinforcer than *right* is a positive reinforcer. Upon the discontinuation of all reinforcement it can be concluded that the combination which during acquisition does not include a positive reinforcer (NW) produces the least amount of extinction. The combination which during acquisition includes a positive reinforcer but not a negative reinforcer (RN, BN) produces the greatest amount of extinction. The combination including both a positive and a negative reinforcer (RW), which presumably permits the original strength of *nothing* to operate, leads to extinction which is greater than NW but not as great as RN or BN. The foregoing generalizations were found at both developmental levels.

The results of this and the earlier study have been interpreted from the point of view that *right* and *wrong* are at opposite ends of the reinforcement continuum with *nothing*, assumed to have little or no initial reinforcement value, placed closer to the weaker reinforcer of *right*. We have further taken

the position that *nothing,* when paired with *wrong* or *right,* acquires, respectively, positive or negative reinforcement value. However, the writers are not asserting that our position is the only one which explains our data. It can be argued, for example, that the extinction data result merely from the extreme change in conditions from acquisition to extinction for the RW, RN, and BN groups whereas the extinction condition for the NW group remains very similar to the acquisition condition. Though such an explanation makes sense, we would assert that such a position limits the possibilities for further research. By maintaining the theoretical properties ascribed to the *nothing* condition, several definitive experiments are suggested [*see* Buchwald (2), for an example]. These experiments may fail to support our position, but, at least, they will have the value of increasing our understanding of reinforcement effectiveness amongst children.

Summary

In a previous experiment involving four reinforcement combinations (Right-Wrong, Right-Nothing, Nothing-Wrong, and Buzzer-Nothing) and two developmental levels (four- and five-year-olds and eight- and nine-year-olds), a significant age $\times$ treatments interaction was reported. Since multiple comparison tests failed to yield significant results, it was decided to replicate the study using those treatments which seemed to contribute most to the interaction terms (RW, NW, BN). Two separate experiments were conducted using age groups and procedures identical to the original study. Comparison of the RW-NW combinations and the NW-BN combinations constituted the first and second experiments, respectively. The results of the experiments failed to replicate the age $\times$ treatment interaction. No difference was found between the RW-NW treatments though the NW group was reliably superior to the BN group. The extinction data replicated in entirety. Reasons for the lack of agreement for the acquisition results were examined. On the basis of the two studies the conclusion was reached that *wrong* is a stronger reinforcer than *right* and that this difference appears as early as five years of age. It was also concluded that regardless of developmental level *nothing* when paired with *wrong* acquires positive reinforcement value and serves to maintain performance at a pre-extinction rate. When *nothing* is paired with *right,* it acquires negative reinforcement value and serves to extinguish behavior.

References

1. Brackbill, Y., & O'Hara, J. The relative effectiveness of reward and punishment for discrimination learning in children. *J. comp. physiol. Psychol.,* 1958, 6, 747-751.

2. Buchwald, A. M. Experimental alterations in the effectiveness of verbal reinforcement combinations. *J. exp. Psychol.*, 1959, 57, 351-361.

3. Buchwald, A. M. Extinction after acquisition under different verbal reinforcement combinations. *J. exp. Psychol.*, 1959, 57, 43-48.

4. Buss, A. H., Braden, W., Orgel, A., & Buss, E. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.*, 1956, 52, 288-295.

5. Buss, A. H., & Buss, E. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, 52, 283-287.

6. Edwards, A. L. *Experimental design in psychological research*. Rinehart, 1950.

7. Meyer, W. J., & Seidman, S. B. Age differences in the effectiveness of different reinforcement combinations on the acquisition and extinction of a simple concept learning problem. *Child Develpm.*, 1960, 31, 419-429.

NORMATIVE DATA ON THE CHILDREN'S MANIFEST ANXIETY SCALE AT THE RURAL THIRD GRADE LEVEL[1]

HAROLD D. HOLLOWAY*
University of Tennessee

The children's form of the manifest anxiety scale (CMAS) developed recently by Castaneda, McCandless, and Palermo (1) has become a valuable research instrument at the fourth, fifth, and sixth grade levels. The predictive merit and reliability of the CMAS using these mid-grade subjects (*Ss*) may be confirmed by referring to the series of articles, given in the references, written jointly by the three above authors and singly by Palermo (7). Holloway (3) presented evidence that the CMAS could also be used reliably (retest $r = .83$) at the rural third grade level.

The purpose of the present study was to furnish additional information concerning the CMAS in the form of normative statistics at the rural third grade level—an essential complement to any newly devised instrument.

METHOD

Subjects

The sample was actually a byproduct of a larger study (hereafter "main study") concerning relations between parental attitudes toward childrearing and the CMAS in Tennessee. In order to obtain CMAS protocols designated by the sampling design of the main study,[2] it was necessary to test all pupils

* Department of Child Development and Family Relationships, University of Tennessee, Knoxville.

[1] The study was sponsored jointly by the Agricultural Experiment Station and the College of Home Economics, Department of Child Development and Family Relationships, of the University of Tennessee. The author wishes to express his gratitude to the public school officials and teachers who participated and granted permission to conduct the study. The author also acknowledges his gratitude to Dr. Mary E. Keister, University of Tennessee, and to Dr. A. L. Finkner, Department of Experimental Statistics, North Carolina State College, who participated as consultants.

[2] The design for the main study was a one-stage stratified random sample, the strata being the three principal geographical regions of Tennessee—East, Middle, and West. The basic sampling unit was a cluster of approximately 15 third graders from a given rural school.

in a given third even though in many instances but a few of the pupils in the room were slated for use in the main study. The procedure resulted in additional usable CMAS protocols. The total number obtained was 524 with approximately 200 of them to be used in the main study. The sample of the present study was constructed from among the 524 protocols simply by equalizing, for analytical reasons, the number of boys and girls by a random procedure within each of three main geographical strata of Tennessee, namely, East, Middle, and West. The resultant number of *Ss* was 462 distributed with respect to sex, number, age, and location as shown in Table 1. The sample cannot be described in the strict sense as being a so-called "representative" or "random" one of the state. However, the results to be reported will constitute the best estimate one can make in such a direction at the present time.

TABLE 1

SEX, NUMBER, AGE, AND LOCATION CHARACTERISTICS OF THE SUBJECTS

	Sex and Number			CA in Years			
Stratum	*Boys*	*Girls*	*Total*	*Mean*	*Range*	*No. of Counties*	*No. of Communities*
East	103	103	206	9.0	7.7–15.2	6	10
Middle	87	87	174	8.8	7.8–12.0	5	5
West	41	41	82	8.9	8.1–11.0	3	3
Combined ..	231	231	462	8.9	7.7–15.2	14	18

The subjects were white, third graders, attending schools in rural areas ("rural" as defined by the 1950 U. S. Census, i.e., open country or places with less than 2500 population—the four main metropolitan counties were eliminated) and attending schools having more than one teacher.

Procedure

The CMAS was administered orally to each classroom by the teacher who followed a set of standard instructions prepared by the present author. The testing program was accomplished during 1957 and 1958.

Results and Discussion

The obtained A scale (42-item "anxiety" subscale of the CMAS) and L scale (11-item "tendency to falsify responses to A" subscale) scores were entered into separate Stratum × Sex analysis of variance designs.[3] The basic means and SDs of the two subscales are presented in Table 2 combined

[3] For a complete description of the CMAS and its A and L subscales, see Castaneda, McCandless, and Palermo (1).

TABLE 2

STRATUM × SEX MEANS AND STANDARD DEVIATIONS FOR A AND L SCALES SEPARATELY

	BOYS		GIRLS		COMBINED SEXES	
Stratum	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
A Scale						
East	22.27	7.28	23.79	7.78	23.03	7.57
Middle	20.05	8.40	20.77	8.05	20.41	8.24
West	19.66	9.58	19.78	9.33	19.72	9.46
Combined Strata	20.97	8.24	21.94	8.35	21.45	8.31
L Scale						
East	5.37	1.59	5.94	1.80	5.66	1.72
Middle	5.28	1.84	5.59	1.75	5.43	1.80
West	5.68	1.69	5.73	1.71	5.71	1.70
Combined Strata	5.39	1.71	5.77	1.77	5.58	1.75

with the total sample statistics. The total sample values will be considered first to be followed by separate presentations of the analyses of variance and the correlations between the A and L scales.

Total Sample Statistics

The A scale mean and SD for the total sample of 462 *Ss* are given in the upper-right section of Table 2. The median (21.98), eightieth and twentieth percentiles ($P_{80} = 29.30$ and $P_{20} = 14.04$, respectively) are reproduced in Figure 1 which is an over-all frequency polygon of the A scale scores. In general, the separate sex and stratum distributions were similar to that shown in Figure 1, that is, approximately bell shaped and unimodal. For girls alone the median was 22.67; P_{80} was 29.84, and P_{20} was 14.90. The corresponding values for boys were: 21.40, 28.47, and 13.10.

Inspection of the frequency distributions plotted similarly for the total sample of L scale scores revealed them also to be approximately bell shaped and unimodal. The mean and SD of the 462 L scale scores are given in the lower-right section of Table 2; the median was 5.74; P_{80} was 7.15, and P_{20} was 4.02. The same statistics for girls were: 5.92, 7.32, and 4.24; for boys, 5.56, 6.96, and 3.86.

Analyses of Variance

Summaries of the two Stratum × Sex analyses of variance, one for the A scale and one for the L scale, have been combined and presented in Table 3.

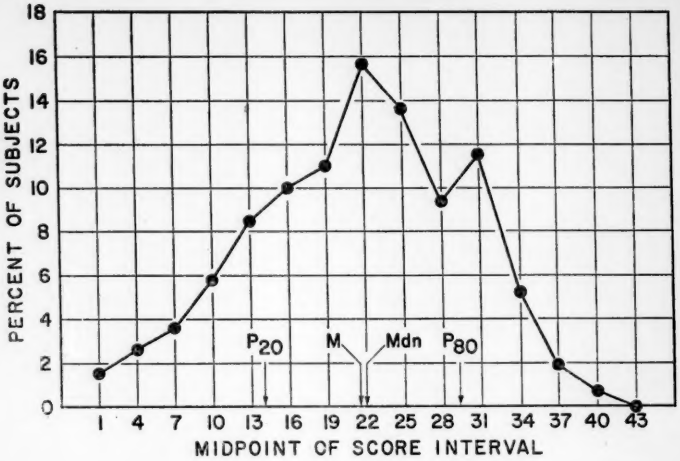

FIGURE 1—Frequency polygon showing percentage of 462 third grade rural children receiving the indicated scores on the A scale of the CMAS.

A scale. It may be observed from the A scale section of Table 3 that the one statistically significant *F* ratio (7.02) indicated an over-all difference among the stratum means. Inspection of these three means in the upper right portion of Table 2 revealed a consistent increase from West (19.72), through Middle (20.41), to East Tennessee (23.03). By definition, the

TABLE 3

STRATUM × SEX ANALYSIS OF VARIANCE SUMMARIES FOR A AND L SCALES SEPARATELY

		A SCALE		L SCALE	
Source of Variation	*df*	*Mean Square*	*F*	*Mean Square*	*F*
Strata	2	474.07	7.02**	3.18	1.04
Sex	1	108.61	1.61	16.76	5.50*
Interaction	2	16.32	< 1.00	2.19	< 1.00
Within-Cells	456	67.53		3.05	
Total	461	69.16		3.08	

NOTE.—Coefficient of risk of $p = .05$ used for all *F* tests.
* $p < .05$.
** $p < .01$.

higher the A scale score, the higher the anxiety. Individual t tests[4] between pairs of means showed that the East mean was significantly higher than either the Middle ($t = 3.08$; $p < .01$) or West ($t = 3.08$; $p < .01$) means with no significant difference between the latter two strata ($t < 1.00$).

Generally speaking, the means were higher than those found in similar studies involving fourth, fifth, and sixth graders in different parts of the country. Whether or not such differences can be attributed to variation in grade level and/or geographical location remains unanswered. The trend for the A scale means to increase from West to East was perhaps the most intriguing result, and one can only speculate as to the reasons. The most obvious possibility (or strata correlate) is the fact that population size, like the anxiety means, increases from West to East. Secondly, the terrain ranges from relatively flat farming land in the West, through semimountainous regions in Middle, to mountainous sections in the East. Sociological experts of the state would certainly posit general cultural differences between peoples living in rural mountain communities versus those in rural nonmountain communities. The author would hypothesize that among such cultural differences are such factors as conservatism and restrictiveness in social philosophy and behavior. It would seem that these two characteristics are weighted more heavily in the East than in the West which, on a speculative basis, is perhaps reflected in increasing anxiety from West to East.

L scale. The analysis of variance yielded a significant F ratio for the sex factor ($F = 5.50$—Table 3), indicating that the mean for girls of 5.77 was significantly higher than the mean of 5.39 for boys (*see* Table 2). The finding was consistent with the research literature involving the CMAS.

A and L Scale Correlations

Ideally, from the standpoint of measurement, the correlation between the anxiety portion of the CMAS (A scale) and the "tendency to falsify" the anxiety items (L scale) should be zero. Pearson rs were computed between A and L for each stratum-sex combination to ascertain an estimate of the relationship. The obtained rs ranged between $-.23$ and .00, none being significantly different from zero using a coefficient of risk of $p = .05$. These correlations were typical of those found in other investigations.

Summary

The purpose of the study was to obtain normative data on the CMAS at the rural third grade level. The CMAS was administered orally by the teachers of 462 third grade rural children, who were equally divided with respect to sex and who ranged in age from 7.7 to 15.2 years with a mean of 8.9 years. The *S*s were drawn from 18 rural communities distributed throughout East, Middle, and West Tennessee. Data were analyzed in terms

[4] Within-cells mean square used as a common estimate of error variance.

of frequency distributions, Stratum × Sex analyses of variance, and correlations. The main results were:

1. The principal A and L scale frequency distributions were approximately bell shaped and unimodal.

2. The A scale means increased consistently from West to East Tennessee with a significant difference occurring between East and each of the other two strata. Sex differences and their interaction with strata were not significant.

3. The L scale mean for girls was significantly higher than for boys. Stratum differences and the stratum-sex interaction were not significant.

4. Pearson *rs* between the A and L scales were not significantly different from zero.

References

1. Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, 27, 317-326.
2. Castaneda, A., Palermo, D. S., & McCandless, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.*, 1956, 27, 327-332.
3. Holloway, H. D. Reliability of the children's manifest anxiety scale at the rural third grade level. *J. educ. Psychol.*, 1958, 49, 193-196.
4. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Houghton Mifflin, 1953.
5. McCandless, B. R., & Castaneda, A. Anxiety in children, school achievement and intelligence. *Child Develpm.* 1956, 27, 378-382.
6. McCandless, B. R., Castaneda, A., & Palermo, D. S. Anxiety in children and social status. *Child Develpm.*, 1956, 27, 385-391.
7. Palermo, D. S. Racial comparisons and additional normative data on the children's manifest anxiety scale. *Child Develpm.*, 1959, 30, 53-57.
8. Palermo, D. S., Castaneda, A., & McCandless, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.*, 1956, 27, 333-337.

SOCIAL HISTORIES OF SUCCESSFUL AND UNSUCCESSFUL CHILDREN

JOHN W. BAKER* *and* ANNETTE HOLZWORTH[1]
Mental Health Research Institute, Fort Steilacoom, Washington

The social histories of children in a state mental hospital and of successful children in a matched control group will be statistically analyzed and compared in this paper.

Information obtained from social histories has long been used in the diagnosis of children's emotional and social problems as well as in the formulation of treatment plans. In more recent years, the social history has been a part of research studies relating to children's problems. Although isolated items found in social histories have been statistically analyzed, the statistical analysis of complete social histories has been infrequent. Obviously, many items in a social history do not lend themselves to statistical analysis. In addition, the social situation of a family or individual is not accurately or completely portrayed in statistical studies. However, because of the widespread use of the social history as a diagnostic and research instrument, a statistical examination of the social histories of various groups is pertinent. A review of the literature did not reveal any studies that attempted a statistical comparison of the complete social histories of successful school children and of children admitted to a psychiatric hospital, two groups that differ greatly in their social adjustment. It is for these reasons that this study should be of interest to professional people who work with emotionally disturbed children.

METHODOLOGY

Seventy-one children between the ages of 13 and 16 years comprise each of the two groups, making a total of 142 subjects included in this study. The social work unit of the Mental Health Research Institute obtained the social histories of these children between the years 1955 and 1958. The two

* Supervisor, Psychiatric Social Services, Division of Mental Health, Department of Institutions, State of Washington, P.O. Box 867, Olympia.

[1] Former Research Assistant, Mental Health Research Institute, Fort Steilacoom, Washington.

Child Develpm., 1961, 32, 135-149.

groups were matched for age and sex. Eighty-four of the subjects were matched with less than two months difference in date of birth, with the largest gap being four and one-half months. The patients were children admitted to Western State Hospital, Fort Steilacoom, Washington during the years of the study. The control group of successful children met the following criteria: (a) they were in the appropriate school grade, and (b) they had a history of good citizenship in school. This control group was chosen at random by selecting subjects at regular intervals from a larger list of successful students in two public schools in Tacoma, Washington. School officials—teachers, guidance department personnel, and principals—selected the original list of students. Therefore, the control group was not just a random sample of children in two public schools but were successful students by school standards. The two schools were chosen from an area thought to be somewhat similar to the patient group in socioeconomic status.

Each group consisted of 35 boys and 36 girls with a total of 70 boys and 72 girls in the study population. In the patient group there were two Negroes and two with mixed racial backgrounds, while the controls were all Caucasian. Although the psychiatric diagnoses of the patient group are not considered in the findings of this paper, they are as follows: schizophrenic—11, schizophrenic with indications of organicity—17, primary neurotic—15, neurotic with a questionable thought disturbance—13, and neurotic with psychopathic manifestations—15. The diagnoses were made by the psychiatric staff of the Mental Health Research Institute.

A psychiatric social worker obtained the social histories from the parents of the control group during an interview lasting approximately two hours in either the parents' home or at the school. The same social history outline was used with the parents of the patient group. When reports from other agencies were available, they were used to supplement the patients' histories. Several different interviewers participated in the study and attempted to obtain as detailed a history as possible in the allotted time.

The outline used in taking the histories covered subjects that are generally used in social histories and perhaps in more detail than usual in a service agency. The parents were asked about the child's prenatal, birth and neonatal history, including signs of early psychosocial difficulties. Developmental, health, and school history was asked, as well as the child's relations with his parents, siblings, and peers. Information was also obtained about the parents' backgrounds and their relations with each other and in the community. An attempt was made to learn of the child's behavior problems or any placements outside the family.

Although six hundred factors were originally listed, because of the inadequate information in some of the histories and the ambiguity of some of the items, only 83 factors were included in the final analysis. The phi coefficient, a measure of the degree of association between two dichotomous variables, was used for the statistical analysis. The lowest level of significance accepted was at the 5 per cent level.

Results

Of the 83 items that were statistically analyzed, 47 were significant at the 1 per cent level and five at the 5 per cent level. These findings, as well as the statistically insignificant ones, will be discussed as far as possible in the chronological sequence of the children's development.

Prenatal and Birth History

The mothers' health during pregnancy was statistically different in the two groups, with 98 per cent of the control mothers reporting good health with no unusual complications during their pregnancies. This compared with 82 per cent of the patient mothers. Poor health ranged from prolonged nausea and vomiting to accidents, injuries, and attempted abortions. The medical history of the mothers during pregnancy was not verified by medical records or reports; however, according to the mothers' perceptions of their pregnancies, significantly more patient mothers felt that their pregnancies were characterized by ill health and unpleasantness.

Other researchers and clinicians have felt that the health and attitude of the mother during pregnancy are closely related and have a direct bearing on the early mother-child relationship. Despres found that "those (mothers) who were not happy with the fact of their pregnancy were significantly more apt to experience nausea and vomiting (and to experience nausea over a longer period of time) than women more favorably disposed toward their pregnancy" (3, p. 617). Josselyn wrote, "the emotional experiences of the mother during pregnancy have considerable bearing upon the attitude of the mother toward the newborn child, and thus upon the early relationship between the child and the mother. If she has had a happy pregnancy the baby enters a world that is, to a large measure, prepared to welcome it happily" (8, p. 28). This study did not attempt to uncover the reasons for the differences in the mothers' reports of their pregnancies. The majority of mothers in both groups, however, reported satisfactory pregnancies. From the available data, differences were not found to indicate that one group had any more abnormal or physically traumatic deliveries, although the information in this area was inadequate.

Twelve of the 71 patients (17 per cent) were born illegitimately, while none of the controls were reported born out of wedlock. Since the control group was an unusually successful sample, the patient illegitimate birth rate can be compared with a more random sample. The national illegitimate live birth rate was 4.2 per cent of all registered births in 1945 (14). The children in this study were either born in 1945 or within two years of this date. As can be seen from these figures, illegitimacy in the patient population is significantly higher than in the control group and in the national population.

The age of the parents at the time of the child's birth was not found to be statistically different in the two groups. The mean age for patient

mothers was 25.3, control mothers, 26.1, and for both patient and control fathers, 29.4.

Developmental History

The age when the child began to talk in sentences was the only factor in this area that was significantly different from the available data. The controls talked in sentences an average of 1.8 months earlier (24.0 months to 25.8 months).

Differences in the average ages at time of walking, talking, and toilet training were insignificant between the two groups and within the norms suggested by Gesell and his associates (4).

Breast feeding has been thought by some to be preferable to bottle feeding for the healthiest psychosocial development of the child (9). This study did not find any significant differences in this area. About half of each group were breast fed; 48 per cent of the patients and 43 per cent of the controls were bottle fed from birth. The incidence of feeding problems in both groups was low and could not be related to either breast or bottle feeding.

Early Psychosocial Difficulties

Of interest to those concerned with the early detection and treatment of emotional and social maladjustments are the findings of early psychosocial problems in the children, as shown in Table 1.

TABLE 1

INCIDENCE OF EARLY PSYCHOSOCIAL PROBLEMS

Type of Problem	PER CENT *Patients*	PER CENT *Controls*
Enuresis*	44	14
Severe temper tantrums	35	13
Fears or phobias	25	..
Speech difficulties	23	4
Rocking or head banging	16	1
Encopresis	8	..

* Beyond 3 years of age.

With one exception—a boy who was enuretic until age 12—the controls' enuresis had stopped by school age. Parents are sometimes told not to be too concerned about problems such as enuresis, exaggerated fears and phobias, etc., as the child will "out grow" them. This is true in some cases as evidenced by the early presence of such problems among the control group; however, these findings give evidence that there is a very significant relation between these early difficulties and the later development of serious psychosocial problems.

TABLE 2

NUMBER OF EARLY PSYCHOSOCIAL PROBLEMS

	NUMBER	
Number of Problems	*Patients*	*Controls*
1	21	18
2	16	..
3	8	1
4	..	..
5	2	..

Although some of the controls had early signs of psychosocial problems, only one had more than one such symptom. This is in contrast to the patients, 36 per cent of whom had a combination of at least two problems, as can be seen in Table 2.

The only significant cross-correlation of these early symptoms was between enuresis and severe temper tantrums. It has been suggested by several people that enuresis beyond the age of 3 years is an expression of hostility or aggression on the part of the child. These findings indicate that, if this is true, the enuretic child is likely to demonstrate his hostility and aggression by other means, particularly temper tantrums, as well. According to Allen in a study of aggression, "the boy with long standing enuresis who said, 'I want to be dry, but I don't want to be made dry,' stated the common dilemma that emerges in the struggle between the parent requiring bladder control and the child fighting against it. Enuretic behavior offers such a safe way to be aggressively oriented against responsibility and separation without the more open danger of saying, 'I will not,' to a power the child feels to be superior" (1, p. 360). If prolonged enuresis is an expression of hostility, the children in this study used temper tantrums as well as enuretic behavior to say "I will not."

The significantly high percentage of the patient group who had early rocking or head banging corroborates the theory that such behavior in the child is an early indication of the possibility of serious emotional problems in the child. Concerning rocking, Spitz has said, "Inconsistent, contradictory behavior of the mother makes the establishment of adequate object relations impossible and arrests the child at the level of primary narcissistic discharge of its libidinal drive in the form of rocking" (10, p. 111). This study did not attempt to determine the causes of rocking or head banging; therefore, the causes as stated by Spitz can neither be confirmed nor refuted.

Almost one-fourth of the patient group had early speech difficulties, such as stammering and stuttering. Another study found "a definite correlation between the development of talking and the adequacy of early care; not only did the well-mothered babies vocalize sooner but their speech

maintained a smoother and easier progress" (9, p. 20). The present study did not reveal a statistically significant difference in the age at which the two groups began to talk, although the control group did begin to talk at a slightly earlier age. However, the large number of patients with speech problems was highly significant. From this study, a cause-and-effect relation cannot be established between severe emotional problems and early speech difficulties but a concomitant association is obvious.

School Problems

Because of the selection of the control subjects, they obviously did not have any serious school problems; however, school was the area of poorest performance for the patients. Academic problems in the first two grades were reported for 66 per cent of the patients; in the later grades this increased to 82 per cent. Again, this may be an important area for the early detection and treatment of problems that later develop into serious psychosocial maladjustments. This is particularly true if early school problems are accompanied by other difficulties. Academic difficulty in the first two grades was highly correlated with enuresis, speech difficulties, rocking or head banging, temper tantrums, overt aggression, and exaggerated fears or phobias.

To illustrate further the early school problems of the patients, 10 per cent were refused further admission to their schools during the first two grades, and an additional 31 per cent were refused in the later grades. Social promotions were given to 18 per cent of the patients. Truancy was a problem in 41 per cent of the patients; the difference is significant at the 5 per cent level in the first two grades and the 1 per cent level in the later grades.

Peer Relations

The ability to make friends and to get along with one's peers is considered to be healthy, normal, and desirable by most people in our society. According to their parents, the controls had no difficulties in this area; all had at least a few good friends and had no serious problems in getting along with others. The histories of the patient group showed that only 34 per cent of the patients had desirable peer relations. The desirability of the relations with their peers was decided by clinical evaluation from information obtained from parents, school, and other agency reports.

Part of their socialization pattern which could be more easily tabulated was group membership. Again, the controls made a far more satisfactory social adjustment than the patients. Each control belonged to an average of 1.7 groups while the patients' average was 0.3. Membership in one or more groups was held by almost three-fourths of the controls and only one-fifth of the patients. In other words, 80 per cent of the patients, in contrast to 27 per cent of the controls, were not members of any group. As another possible indication of the socialization patterns, 36 per cent of the patients and 86 per cent of the controls attended church at least once a month.

A study of a rural teenage population in the State of Washington found that 80 per cent of the subjects belonged to at least one social group (11). This is slightly more than the control group (73 per cent) and considerably more than the patient group (20 per cent) in this study. All of the controls lived in an urban area; however, the patients were both rural and urban. From these figures, it appears that the higher percentage of urban children in the control group does not in itself account for the differences in group memberships in the two groups.

Delinquent Behavior

Because of the bias in the selection of the control group, one would not expect to find a history of blatant delinquent behavior in the controls. Only one of the control children was reported by his parents to have been involved in delinquent activities; he was involved in a minor theft and spent one night in the juvenile detention home before being released. Forty-six (65 per cent) of the patients had been involved in delinquencies. It is interesting that more than half (27 of 46) of the delinquents committed their delinquent acts with other children rather than alone.

Of these patients with delinquency histories, 52 per cent were involved in runaways from home; 44 per cent, thefts (other than car thefts); 42 per cent, overt aggressive and assaultive behavior; 27 per cent, sexual deviancy; 22 per cent, auto theft; 18 per cent, vandalism; and 13 per cent, firesetting.

Placement Outside the Home

An overwhelming majority of the patients (72 per cent) had been placed outside their own homes prior to hospitalization. Only three of the controls had lived with anyone other than their immediate families, all three having lived with other relatives.

The patients had an average of 3.3 placements before coming to the hospital: 55 per cent had previous placements in institutions, 27 per cent with relatives, and 24 per cent in foster homes; therefore, a large number of the patients had been tried in other institutions before their hospitalization. The average length of their placements was 3.8 years. Thirty-two of these placements lasted only one year or less, while at the other extreme, two lasted for more than 10 years.

The average age of the patients at the time of first placement was 8.2 years, ranging from birth to 14 years of age. Approximately half of the first placements were caused by delinquent and incorrigible behavior on the part of the children and the other half by family break-up and parental neglect.

Medical History

The items in the medical and health histories of the children were excluded from the final analysis because of inadequate or ambiguous information. However, a history of known and reported head injuries was approximately the same in the two groups. Minor head injuries without loss of con-

sciousness and/or hospitalization were reported in the histories of 15 per cent of the controls and 20 per cent of the patients. Head injuries with loss of consciousness and/or hospitalization were found in the histories of 15 per cent of controls and 18 per cent of the patients. Statistically insignificant was the incidence of epilepsy in the two groups.

Family Composition

Much has been written about the impact of a broken home on children's emotional and social well-being. Only 35 per cent of the patients belonged to intact families in which the child was living with both natural parents, in sharp contrast to 86 per cent of the controls. In other words, 65 per cent of the patients and only 14 per cent of the controls came from broken homes. Thirteen patients and four controls had stepparents, and the remaining five patients and three controls were living with their adoptive parents.

In addition, the marriages of the controls' natural parents had lasted longer than had the marriages of the patients' natural parents. The mean duration of the natural parents' marriage was 18.8 years for the controls and 12.0 for the patients.

The age of the child at the time of the break-up of his natural family is of interest. The patients had a mean age of 4.4 years when they were first separated from one or both of their natural parents, approximately the same as in the control group. Of the nine control children from broken families, five were over 4 years of age at the time of the break-up; the other four were 4 years of age or younger.

The number of patients from broken homes was considerably higher than those found in a study of less seriously disturbed subjects. Ingham, in his study of out-patient neurotics, found that 31.1 per cent of the neurotic patients came from homes broken by separation and 1.8 per cent from homes broken by death. The total 32.9 per cent had their homes broken prior to their twenty-second birthday, an age considerably older than the mean age of 4.4 for the patients in this study. A study done in the State of Washington in nine rural public high schools (13) found that 80.7 per cent of the students came from intact homes, a figure just slightly lower than this study's urban control group.

The size of the family in the patient group was significantly larger. The families of the patients had an average of 4.0 children compared to 2.8 in the families of the controls. Hawkes, Burchinal, and Gardner concluded ". . . that a small family environment does not have a detrimental effect on the children's personality development. . . . there was a suggestion that children from smaller families might fare better psychologically than children from larger families. Certainly studies employing a larger sample or ones in which important variables such as socio-economic status and perhaps age differences among the children have been controlled need to be initiated" (6, p. 68). The sample in this study was not larger than the Iowa study but age was controlled and socioeconomic status was approximated. The control

children, the group that came from smaller families, obviously "fared better psychologically" since the patient group was seriously enough disturbed to warrant psychiatric hospitalization.

The ordinal position of the children in the two groups was not significant. Six in each group were the only child in their families. An interesting but not significant comparison of the two groups was that more patients than controls were middle children, ranging from next to the oldest to next to the youngest. Thirty-five controls and 25 patients were the oldest children in their families, and 17 controls and 11 patients were the youngest.

Mothers' History

Several significant items differentiated the patients' and controls' mothers. The following problem areas were found in the histories of the patients' mothers: hospitalization for mental illness, 7 per cent; epilepsy in the immediate family, 10 per cent; excessive drinking, 15 per cent; and chronic physical illness with or without hospitalization, 10 per cent. None of the controls' mothers reported these problems.

The control mothers had a higher level of education, 11.5 to 10.7 years. In the control group 48 mothers had completed the twelfth grade and 21 of these had some college or vocational training. On the other hand, only 32 of the patients' mothers had completed the twelfth grade and only seven had additional education or training.

There was no statistically significant difference between the ages of the mothers at the time of their marriage or at the time of the birth of the child studied. The average age at marriage was 21 years for both groups, and the age at the time of the child's birth has been given in a previous section of the paper. The average number of marriages did not prove to be statistically significant, 1.1 for the control mothers and 1.7 for the patient mothers.

Working mothers have been the topic of many discussions in both professional and popular literature. Many mothers who want or have to work outside the home have felt guilty because of doing so. In this study only slightly more of the patients' mothers worked outside the home, 33 to 25. However, the patients' mothers earned significantly more income per month, $202 compared with $181. Since the controls' mothers had more education and training, speculation might attribute the higher income of the patients' mothers to longer working hours. Of interest in these findings is a conclusion of the Glueck and Glueck report: "Equal proportions of mothers of nondelinquents and delinquents were regularly employed but a greater proportion of mothers of delinquents than nondelinquents worked irregularly" (5, p. 330). From the available data in this study, the regularity of the mothers' employment could not be determined.

The history of mental deficiency in the mothers and their families failed to show any statistical significance as did the mothers' criminal or court records.

Fathers' History

One of the significant differences in the two groups of fathers was their educational attainment. As with the mothers, the controls' fathers averaged more education, 11.5 to 10.5 years. In the control group 45 fathers had completed high school and 15 had further education or training. In the patient group the figures were 23 and 10 respectively. The controls' fathers reported no history of problems with alcohol or arrests. In contrast, the histories of the patients' fathers revealed that 14 per cent had a criminal or court record and 18 per cent had a history of excessive drinking.

The mean occupational level of the two groups of fathers was not statistically different. Most of the fathers in both groups worked as operatives or craftsmen. However, certain differences in the job classifications were found. Grouping the fathers by the U. S. Census Bureau occupational classifications, the frequency of the patient fathers in the extreme categories, professional-managerial and service-labor, was significant. The loading of the patients' fathers in the extreme categories may be due to the classification system used. For instance, included in the professional group of the patients' fathers was a jockey with an unstable work history.

The control fathers had a significantly higher income per month, $464 to $348. A more stable work history for the control fathers was revealed in the findings and may account for the higher income in spite of the similarities of the mean occupational levels of the two groups.

Approximately the same number of control fathers (44 per cent) and patient fathers (37 per cent) had been in the armed forces. The available data did not provide more detailed information about their service careers. In addition, the two groups of fathers were not significantly differentiated by the following items: hospitalization for mental illness, epilepsy in their immediate families, a history of mental deficiency, number of marriages, and age at time of marriage. The average number of marriages was 1.3 for the patients' fathers and 1.1 for the controls' fathers. There was only a one year difference in their average ages at time of marriage: patient fathers, 25.6; control fathers, 24.6.

Family Socioeconomic Status

The data did not lend itself to the establishment of socioeconomic status by the use of such criteria as established by Warner, Meeker, and Eells (12). However, certain conclusions can be drawn from the available data. Not only did the control fathers earn more than the patient fathers, but also the control families' total family income was significantly higher. The mean family income for the control group was $513 per month and exceeded the patient group by exactly $100. The mean family income for both groups was higher than that for the nation in 1955, which was $387 per month (14). Although the patients' mothers earned more than the controls' mothers, the difference in the fathers' income was sufficient to give the controls a higher family income. Although both groups had a family income higher than the

national average, the control families had a higher income with fewer children to support.

It has already been mentioned that the mean occupational level of the fathers was similar. Although Warner found that educational level has a very low correlation with other socioeconomic indices, other studies have used educational attainment as a criterion for establishing socioeconomic status (2). As had already been reported, the educational level of both parents was significantly higher in the control group of this study.

At the time the histories were taken, 20 per cent of the patient families and none of the control families were receiving financial assistance from a public agency. The assistance received by the patients' families was usually from the Aid to Dependent Children program. An additional 10 per cent of the patient group and 1.4 per cent of the control group had received public assistance in the past but were self-supporting at the time of the study.

Since many of the patients came from families with multiple problems, it was not surprising that 30 per cent had a history of financial problems severe enough to warrant public assistance. Perhaps more surprising is that one-third of the patient families who had ever received public assistance were self-supporting at the time of the study. Equally surprising may be the fact that one of the successful control children belonged to a family that had received public assistance in the past.

Integration of Families in the Community

Participation in community organizations clearly distinguishes the patient and control families. Of the control families, 87 per cent were actively associated with at least one community organization in contrast to only 22 per cent of the patient families. Church attendance at least once a month was reported by 58 per cent of the control families and 38 per cent of the patient families. Significantly more controls than patients attended church with their parents.

Home ownership also differentiated the two groups: 89 per cent of the control families compared to 47 per cent of the patient families owned or were buying their homes.

Discussion

The limitations of this study should be examined before evaluating the findings. Parents of both groups knew why their children were being studied, and this undoubtedly had an influence on the information that they gave the interviewers. The parents of the patients were well aware that their children were in a state mental hospital and considered unsuccessful by many people. Anyone who has interviewed the parents of emotionally disturbed children should be aware of the feelings, such as guilt, denial, rationalization, and projection, that the parents bring with them to the

interview, as well as the fear of being blamed or criticized. The possible coloring and distortion of factual information by such feelings is well-known to all in the mental health fields.

Although the information given by the patients' parents was often compared with reports from other agencies, this was not possible in all cases. In addition, other agency personnel, sometimes oriented to look for pathology, knew that the child had been admitted to a state mental hospital when their reports were requested. Therefore, it is possible that agency reports emphasized the pathological aspects of the child's history and behavior.

On the other hand, the parents of the control children knew that their children were chosen for study because they were "successful." It is impossible to know the effect of this knowledge on their perceptions and descriptions of the children's histories. Both control parents were not always seen for the history interview; therefore, the recollections and the opinions of just one parent often had to be accepted as valid.

Another factor that affected the study's findings was the inadequacy of information in the patients' records. Not all of the patients' parents were interviewed. Often just one parent gave the social history, usually the mother. The patients' parents were less able to remember and relate parts of the history than were the controls' parents. Since many of the patients had been placed outside their parents' homes, the parents lacked firsthand information about certain items in their histories. In some cases, agency reports were able to fill these gaps.

The interviewers knew that the child was in either the patient or control group before they took the history. This information may have biased the interviewer in the type of information he obtained and the interpretation of the findings. Although the histories of the control subjects were taken by professionally trained psychiatric social workers, the histories of the patients sometimes were taken by less trained personnel.

Therefore, because of a possible bias in the perceptions of both the interviewers and the parents, pathology may have been emphasized in the histories of the patients and normalcy may have been emphasized in the histories of the controls.

Still other chances for human error should be considered. Several different people were involved in the collection and tabulation of the data. While this has its advantage in eliminating the personal bias of just one person, it has its disadvantages in lack of uniformity and continuity.

Although many statistically significant differences were found between the patient and control groups, few of the items were found to occur exclusively in either group. The two groups were selected because of their extreme difference in social adjustment; therefore, significant differences in their social histories would necessarily be expected. The causes of the differences and similarities in the two groups were not investigated in the study. More detailed information would be desirable for several of the items but was beyond the scope of the study.

In spite of its limitations, the study does contribute some reliable findings of interest and significance. It almost goes without saying that many of the areas discussed in this paper need and warrant further investigation. Perhaps one of the greatest values of statistical studies is their ability to indicate areas for more intensive research.

Conclusions

Similarities in the Two Groups

A comparison of the social histories of the patient and control groups reveals some interesting similarities. The parents of both groups were about the same age not only at the time of their marriage but also at the birth of the child studied. A majority of the parents in either group had been married only once. The percentage of working mothers in both groups was approximately the same, and slightly less than half of all the fathers had been in the armed forces.

In addition, approximately half of the children in each group were breast fed and half were bottle fed from birth. Feeding problems were rare in both groups. The average ages at time of walking, talking, and toilet training of all the children were within normal limits and about the same in both groups. Very few of the children were the only child in the family, and their ordinal positions were similar. A very small percentage of the children in each group was adopted. Although the health histories of the children were inadequate, head injuries were reported to about the same degree in the two groups.

The Patient Group

A composite picture of the patient group shows some interesting contrasts to the control group. Although a majority of all the mothers reported good health during their pregnancies, more of the patients' mothers told of having difficult or complicated pregnancies. Over 10 per cent of the patients' births were illegitimate. The ability to talk in sentences came more slowly for the patients. Early psychosocial problems, such as enuresis, severe temper tantrums, fears and phobias, speech difficulties, rocking, and head banging, were prominent during the patients' infancy and early childhood. The children who were enuretic also tended to have severe temper tantrums. Those children who had early psychosocial problems also had difficulty in their school performance. School achievement and adjustment were serious problems for the patient group and appeared in the early years of their school histories.

A large majority of the patients had poor relations with their peers and did not belong to any social group outside their family and classroom. More than half of the patients were involved in delinquent behavior, and they tended to get involved in delinquencies with other children rather than alone. Almost three-fourths of the patients had been placed in foster homes

or other institutions prior to their hospitalization with an average of over three such placements.

An overwhelming majority of the patients came from broken homes. The patients also came from larger families than did the controls. Patients' families had lower family incomes, and their fathers had a more unstable work history. Although only slightly more patients' mothers worked outside the home, they earned significantly more than the controls' mothers. Public assistance was much more often needed by the patients' families. One-third of those receiving such help eventually became self-supporting.

Both parents of the patients had an average of less than an eleventh grade education. The mothers of the patients were more likely to have been in mental hospitals and to have had problems with alcohol. One-third of the patients' mothers had a history of epilepsy in the immediate families. The patients' fathers were differentiated by court and criminal records, as well as excessive drinking. Patient families were not well integrated in the community and took little part in community organizations. They also tended to live in rented family dwellings.

The Control Group

The successful control child was born of married parents, and his mother's pregnancy was healthy and without complications. This child talked in sentences at an earlier age. Early signs of psychosocial problems were an exception rather than a rule. If the child had been enuretic at all, the enuresis stopped before he started to school. The control children had no serious problems in school or in their ability to get along with other children. Participation in social groups outside the family and classroom was common for these children, as was church attendance. Delinquent behavior was virtually unknown among this group. The successful child had never experienced the trauma of living with anyone other than his relatives and was very likely to be living with both natural parents.

The control family was smaller and had a significantly higher family income. A more stable work history and a higher income were found among the control fathers. The need for public assistance was very rare among this group of families. Although the parents of these children averaged less than a high school education, their educational level was higher than the parents of the patients. Family histories of mental illness and problems with alcohol were not reported by the control parents. These parents were active in community and religious organizations and were likely to own their homes or be in the process of buying them.

Summary

The social histories of children in a state mental hospital were compared with those of successful children in a control group, matched for sex and age. A total of 142 children, between the ages of 13 and 16 years, were

included in the study. Of the 83 items in the histories that were statistically analyzed, 52 were significant at or beyond the 5 per cent level. The differences and similarities in the social histories of the patient and control groups were presented and discussed. Some traditionally held theories were confirmed and others were refuted or questioned by the findings of this study.

References

1. Allen, F. Aggression in relation to emotional development, normal and pathological. *Ment. Hyg.*, 1950, 34, 353-363.
2. Burchinal, L., Gardner, B., & Hawkes, G. R. Children's personality adjustment and the socioeconomic status of their families. *J. genet. Psychol.*, 1958, 92, 149-159.
3. Despres, M. Favorable and unfavorable attitudes toward pregnancy in primiparae. Unpublished Ph.D. dissertation, Univer. of Chicago, 1936. (Secondary source: Wallin, P. & Riley, R. P. Reactions of mothers to pregnancy and adjustment of offspring in infancy. *Amer. J. Psychiat.*, 1950, 20, 616-622.)
4. Gesell, A., & Ilg, F. L. *The child from five to ten.* Harper, 1946.
5. Glueck, S, & Glueck, E. Working mothers and delinquency. *Ment. Hyg.*, 1957, 41, 327-352.
6. Hawkes, G. R., Burchinal, L., & Gardner, B. Size of family and adjustment of children. *Marriage Fam. Living*, 1958, 20, 65-68.
7. Ingham, H. V. A statistical study of family relationships in psychoneurosis. *Amer. J. Psychiat.*, 1949, 106, 91-98.
8. Josselyn, I. *Psychosocial development of children.* Family Service Association of America, 1948.
9. Ribble, M. *The rights of infants.* Columbia Univer. Press, 1943.
10. Spitz, R. Autoeroticism: some empirical findings and hypotheses on three of its manifestations in the first year of life. *Psychoanal. Stud. Child*, 1949, 3-4, 85-119.
11. Stone, C. L. A summary of the responses of Ferry County teenagers to a questionnaire on activities and social relationships. Unpublished ms., Washington State Univer., 1958.
12. Warner, W. L., Meeker, M., & Eells, K. *Social class in America.* Science Research Associates, 1949.
13. Responses of Stevens County high school students to a questionnaire on activities and social relationships of teenagers. Unpublished ms., Dept. of Rural Sociology, Washington State Univer., 1958.
14. Statistical Abstract of the United States, 1957. Bureau of the Census, U. S. Dept. of Commerce, U. S. Gov't Printing Office, Washington, D.C., 1957.

CHILD BEHAVIOR PROBLEMS AND PARENTAL ATTITUDES[1]

DONALD R. PETERSON,* WESLEY C. BECKER, DONALD J. SHOEMAKER, ZELLA LURIA,[2] *and* LEO A. HELLMER
University of Illinois

If modern personality theory and clinical practice have any basis in fact, the importance of parental attitudes in the origin of behavior problems among children can be assumed. The nature of the relationships between parent and child behavior, however, cannot merely be assumed, but must either be derived from or confirmed by empirical research. For some time, the investigators have been engaged in a program of study designed to clarify certain associations between parental characteristics and behavior tendencies of children. The work parallels a great deal of previous research in this area, but differs from most of the latter in the systematic inclusion of variables derived from examination of fathers. In most contemporary research, and in the usual clinical management of child behavior problems, emphasis on the maternal influence is overwhelmingly strong. Consequent neglect of the attitudes and other qualities of fathers has, until now, been justified only by assumption or convenience, and the assumption has seemed to us to require test.

Two studies (1, 5), involving the examination of characteristics of both parents in reference to certain behavior tendencies among children, offered such a test and yielded, among others, the following results:

1. Contrary to general assumption, the attitudes of fathers were at least as intimately related as the attitudes of mothers to the occurrence and form of maladjustive tendencies among children.

---

* Department of Psychology, University of Illinois, Urbana.

[1] This study was supported by the Department of Public Welfare, State of Illinois. We extend our gratitude to the participating parents and teachers, especially of Westview and Bottenfield Kindergartens in Champaign, for their part in providing the data, our high praise to Herbert Quay and Rodgers Elliott for their management of the many administrative problems of the study, and our appreciation to Irvin Moelis and Geraldine Piorkowski for their help in the data analysis.

[2] Now at Tufts University.

Child Develpm., 1961, 32, 151-162.

2. Personality problems (shyness, withdrawal, etc.) among children attending a guidance clinic were relatively independent of maternal attitudes, but were related to autocratic attitudes and lack of parental concern among fathers.

3. Conduct problems (truancy, stealing, etc.) were associated with general maladjustment among mothers and with evident permissiveness and disciplinary ineffectuality on the part of fathers.

Like any such findings, these permit certain conclusions to be drawn about parent-child relationships, but generalizations are limited by the specificity of the methods used and the particular samples chosen for investigation. For this reason it is heuristically advantageous to regard such results as guiding hypotheses for further research, and this is what we have done in the present investigation. With improved, but comparable, methods of measuring parental attitudes and child behavior and with differently selected samples of subjects, we have sought to assess the generality of the above propositions and thus to evolve more substantial conclusions about the way parents and children affect each other.

Subjects

The study deals with two groups of families. Children in the first group were drawn from two local kindergartens solely on the basis of willingness to participate in the research. We shall call them the *Kindergarten* sample. Families in the second group had come to the University of Illinois Psycho-

Table 1

SAMPLE CHARACTERISTICS

	Kindergarten Group		Clinic Group	
	Mean	*SD*	*Mean*	*SD*
Age of child (in months)	65.3	2.9	69.8	7.1
Age of father (in years)	35.3	6.0	33.7	4.6
Age of mother (in years)	33.0	5.2	31.5	3.5
Education of father (grades completed)	16.0	2.6	15.1	2.7
Education of mother (grades completed)	14.3	2.0	13.0	1.5
Sex of child:				
Number of boys	24		17	
Number of girls	29		7	
Occupation of father (Warner):				
Level 1 (professional, etc.)	9		5	
Level 2	30		4	
Level 3	10		8	
Level 4	2		2	
Level 5 (skilled labor, etc.)	2		5	

logical Clinic to deal with one kind of child behavior problem or another. Let us call them the *Clinic* sample. An important consideration in forming the latter group was to select children of approximately kindergarten age, but, beyond that, few restrictions were imposed. Cases of obvious intellectual defect and of known organic disability were not included, nor were those in which the parents were separated or unavailable for study. All subjects were white. An effort was made to exclude families involving adoptive relationships of any sort, but one such family was erroneously admitted anyway and retained for subsequent analyses. By the time the mistake was discovered, it was too late to do anything about it. Various characteristics of the samples are described in Table 1.

The Kindergarten sample, particularly, and both samples, to some extent, represent a restricted upper range of the national population in respect to socioeconomic, educational, and intellectual status.[3] The generality of the study is therefore limited, but, to make a virtue of necessity, the reduction in variance along such extraneous dimensions as intelligence should attenuate relationships between them and the factors with which the study is centrally concerned.

In a clinic the size of ours, the number of kindergarten age children admitted in any given year is small, and we were forced to take practically all such clients who came along. As a result, the quality of control leaves something to be desired. The Kindergarten and Clinic groups differ in respect to sex of child (there are over twice as many boys as girls in the Clinic group), age of child (Clinic children are older, on the average, and vary more in age than those in the Kindergarten group), education of mother (Kindergarten mothers went to school longer), and occupation of father (occupational level was higher for fathers in the Kindergarten group). The possible effect of these differences on subsequent comparisons is difficult to assess, but an effort to do so was made by correlating each of the background variables with parental attitudes as we measured them. Only two of the 40 pertinent relations resulting from that analysis were "significant" at the .05 level. The generally low magnitude of the correlations, and the fact that many of the relations were such as to reduce rather than to elevate the effects we shall discuss below, suggests that differences in background were not crucial factors in determining results.

Procedures

The structured interview developed by Sears, Maccoby, and Levin (6) was modified slightly to permit use with fathers and administered very much as Sears *et al.* have described it with mothers. All interviews were recorded, and two judges rated each subject in respect to various parental attitudes and practices. From these ratings, scores were obtained for five of the seven

[3] No measures of intelligence were obtained, but school records showed that the mean IQ for children in the two kindergartens regularly approximates 120.

factors which Sears and his colleagues derived in the course of their study. Unit weights were employed throughout. The elemental variables comprising each factor and the factor designations themselves are set out below:

Factor 1. *Strictness.* Restrictions on play, demands for orderliness, demands in regard to table manners, severity of toilet training, standards of obedience, emphasis on scholastic achievement, (negative) response to dependency, use of physical punishment, restrictions on and punishment for aggression against parents and siblings, restrictions on masturbation, nudity, and sex play.

Factor 2. *Adjustment.* Self-esteem, esteem for spouse, satisfaction with life situation, (low) child-rearing anxiety, delight of self and spouse over pregnancy, (little) disagreement over child-rearing, report of affection between spouse and child, tendency to follow through on disciplinary threats and promises.

Factor 3. *Warmth.* Warmth, acceptance of dependency, use of praise and reasoning in discipline.

Factor 4. *Responsibility* Self-esteem, demands in respect to orderliness and table manners, restrictions on play, use of praise when children play nicely, standards of obedience, deprivation of privileges, use of tangible rewards in discipline.

Factor 5. *Aggression.* Demands that child "fight back," (low) permissiveness and high punishment for aggression against parents, use of physical punishment in discipline.

Independent evaluations of child behavior were obtained from parents and teachers, using a problem checklist developed for use in this investigation. To select a nonarbitrary sample of problems, 427 representatively chosen case folders for children attending the Clinic were inspected, and the referral problems of each child were recorded. Principally on the basis of frequency of occurrence, 58 problems were selected for general investigation. Both parents and two teachers rated each of the Kindergarten children who participated in the study, and both parents and one teacher rated the children in the Clinic group. In addition, the teachers completed checklists for all nonparticipating subjects in one of the kindergartens. The latter data ($N = 62$) were factor analyzed to permit specification of reasonably unitary, independent dimensions of problem behavior. Since three of the original 58 problems (soiling, enuresis, vertigo) were never observed, they were excluded from further consideration. Phi coefficients of intercorrelation were computed for the remaining 55 variables, and 10 factors extracted by the complete centroid method. Since the first three factors accounted for 55 per cent of the common variance represented in the 10 factors examined and none of the rest of the factors accounted for more than 10 per cent of that variance, only the first three were retained for rotation. To exclude any possibility of subjective bias in defining problem dimensions, Kaiser's varimax criterion (4) was employed as the sole basis for determining vector placement. Factor loadings are presented in Table 2. The first two factors seem rather clearly to represent the conduct problem and personality problem dimensions of concern in our previous work and have been named

TABLE 2

ROTATED FACTOR MATRIX

	Factor Loading		
	1	*2*	*3*
1. *Conduct Problem*			
Disruptiveness	70	—17	15
Boisterousness	70	—30	08
Disobedience	68	—01	02
Attention-seeking	66	—26	—05
Uncooperativeness	62	—05	15
Restlessness	60	—18	28
Fighting	59	—09	06
Negativism	59	15	—12
Impertinence	59	04	—23
Irritability	55	04	—16
Destructiveness	52	—12	28
Distractibility	52	16	38
Irresponsibility	51	05	39
Hyperactivity	50	—15	—06
Temper tantrums	49	28	—07
Laziness	41	08	28
Tension	39	28	—11
Profanity	32	—15	07
Dislike for school	30	23	—28
Thumb-sucking	29	—04	15
Headaches	26	26	—17
Jealousy	23	10	—19
Specific fears	—19	00	16
2. *Personality Problem*			
Inferiority feelings	—02	58	27
Tendency to be flustered	—07	58	12
Lack of self-confidence	—09	57	26
Anxiety	06	55	—07
Self-consciousness	—28	52	05
Aloofness	—19	52	—09
Shyness	—32	51	21
Reticence	10	49	14
Social withdrawal	—20	47	08
Inability to have fun	—11	45	02
Hypersensitivity	16	42	—37
Depression	—03	39	07
Lack of interest in environment	05	39	16
Nervousness	35	36	—16
Crying	28	35	—29

(*table continued on next page*)

TABLE 2 (*continued*)

ROTATED FACTOR MATRIX

	Factor Loading		
	1	*2*	*3*
Preference for older playmates	02	—28	—03
Skin allergy	—03	—22	—01
3. *Autism*			
Preoccupation	—04	20	50
Shortness of attention span	29	22	50
Clumsiness	08	27	50
Inattentiveness	39	24	47
Passivity	01	27	46
Daydreaming	03	38	44
Sluggishness	—20	23	40
Stuttering	06	06	39
Hay fever or asthma	08	—02	37
Drowsiness	—07	20	36
Preference for younger playmates	15	—01	36
Abdominal pain	12	16	—25
Nausea	03	02	—25
Masturbation	10	11	—17
Truancy from school	10	11	—17

accordingly. The third factor is more difficult to interpret, but seems to contain a preponderance of autistic elements and has been labeled to conform with that inference. Its meaning is not unlike that of the tentatively identified "prepsychotic" factor isolated by Himmelweit (3) and described by Eysenck (2). Factor scores for the subjects who participated in the research were computed by assigning unit weights to the 10 most heavily saturated variables in each of the three dimensions.

Analysis of relations between parental attitudes and child behavior proceeded in two ways. First, the attitudes of Kindergarten and Clinic parents were compared. Distributions of the five attitude scores were formed separately for the mothers and fathers in the Kindergarten group, and cutting points established as close as possible to the first and third quartiles in each of the distributions. Attitude scores for Clinic parents were then tabulated on the same baseline, and χ^2 tests of significance performed. In the second analysis, correlations between parental attitudes and child problem factors were determined for the Kindergarten and Clinic groups. Three of the 24 Clinic children were not in school, teacher ratings were therefore unobtainable, and the pertinent correlations reported in Table 5 below are based on an *N* of 21 cases.

TABLE 3

INTERJUDGE CORRELATIONS

Parent Attitude	*Mothers*	*Fathers*	*Child Problem*	*Inter-Parent*	*Inter-Teacher*	*Parent-Teacher*
Strictness	.74	.61	Conduct problem .	.48	.82	.41
Adjustment ...	.70	.72	Personality problem	.52	.68	.24
Warmth	.57	.68	Autism	.62	.74	.41
Responsibility .	.59	.45				
Aggression ...	.81	.81				

Interjudge correlations for the various factors are given in Table 3. These were computed for the Kindergarten sample only, since only in that sample were data from two raters available for all measures. With the exception of the teacher ratings for Clinic subjects, all variables used in the correlational analyses represented average ratings of two judges. The *r*s as reported can therefore be corrected by means of the Spearman-Brown formula. When this is done, interjudge consistencies range from .64 for parent ratings of conduct problem to .90 for teacher ratings of the same dimension. The mean corrected *r* over all reliabilities pertinent to the analysis is .80.

Correlations between parent ratings and teacher ratings of child behavior, however, are relatively low. It would have been convenient and parsimonious to construe these problem dimensions as general, trans-situational tendencies, but to do so seemed operationally unwarranted. Ratings by parents and by teachers were therefore considered separately in all analyses.

RESULTS

Results of the comparison of Kindergarten and Clinic parents are given in Table 4. An inordinate share of both mothers and fathers in the Clinic group were described as poorly adjusted and lacking in warmth. In addition, a significant proportion of the Clinic mothers were said to be irresponsible, and a relatively large number of the Clinic fathers were described as strict and aggressive. The composite attitudinal picture for Clinic mothers, therefore, is of a cold, maladjusted woman, evidently unable or unwilling to assume responsibility for the rearing of her children. The modal Clinic father appears to be a cold, maladjusted, punitive man, lacking the human kindness shown by the fathers of the relatively well-adjusted kindergarten children.

Relations between judgments of parent attitudes and child behavior are presented in Table 5. The most striking findings for the Kindergarten

TABLE 4

COMPARISON OF KINDERGARTEN AND CLINIC PARENTS

	MOTHERS			FATHERS		
Parent Attitude	*Kindergarten*	*Clinic*	χ^2	*Kindergarten*	*Clinic*	χ^2
Strictness			.75			6.81*
high	14	6		13	13	
average ..	26	14		26	6	
low	13	4		14	5	
Adjustment ...			8.43*			23.26*
high	12	1		13	1	
average ..	30	11		27	3	
low	11	12		13	20	
Warmth			14.65*			11.67*
high	16	1		14	1	
average ..	23	6		26	8	
low	14	17		13	15	
Responsibility .			14.02*			1.56
high	14	1		15	5	
average ..	23	5		25	15	
low	16	18		13	4	
Aggression ...			.31			9.77*
high	16	7		11	13	
average ..	21	11		31	6	
low	16	6		11	5	

* Significant at or beyond the .05 level.

sample are those which indicate covariation between strictness, lack of warmth, and aggression among fathers and problems of all kinds among the children. Contrary to expectation, the various problem dimensions were not differentially associated with specific attitudinal patterns. Conduct problems, personality problems, and autistic tendencies were all related in much the same way to the cluster of punitive attitudes mentioned above. However, there appeared to be some specificity in regard to the situation in which the problem behavior was observed. Strict, cold paternal attitudes, for instance, were most closely related to personality problems as judged by parents, but were more highly correlated with conduct problems as teachers observed them in school. It is entirely possible that strict, cold fathers simply will not allow emotional display in the home. A considerable amount of tension would probably result from sustained suppression of this

TABLE 5

CORRELATIONS BETWEEN PARENTAL ATTITUDES AND CHILD PROBLEMS
(Decimal Points Omitted)

	MOTHERS						FATHERS					
	Conduct Problems		*Personality Problems*		*Autism*		*Conduct Problems*		*Personality Problems*		*Autism*	
	PR†	TR‡	PR	TR	PR	TR	PR	TR	PR	TR	PR	TR
KINDERGARTEN SAMPLE (*N* = 53)												
Strictness ...	11	09	−13	10	20	13	16	34*	34*	08	50*	16
Adjustment .	04	−12	−18	−20	−06	−14	−17	−15	−26	04	−22	−03
Warmth	05	04	20	−23	−01	−25	−17	−40*	−30*	−15	−24	−08
Responsibility	14	13	13	−12	32*	−01	08	07	26	09	33*	07
Aggression ..	16	14	−14	19	25	27*	11	20	34*	25	40*	17
CLINIC SAMPLE (*N* = 21–24)												
Strictness ...	07	23	−43*	−49*	−25	−29	38	23	−19	−33	04	−02
Adjustment .	−35	−34	−32	−15	−36	−20	−29	−18	−35	−20	−02	00
Warmth	−06	−24	14	30	23	21	−01	16	15	−04	23	24
Responsibility	25	01	−38	−54*	−09	−39	38	29	−34	−50*	15	07
Aggression ..	32	25	−19	−37	05	−17	23	57*	−10	−25	10	10

* Significant at or beyond the .05 level.
† Parent ratings of child problem.
‡ Teacher ratings of child problem.

kind, to find outlet as conduct disturbance in the freer atmosphere of kindergarten.

The situational difference discussed above is an exception rather than a rule, and, in general, correlations between parental attitudes and teacher ratings of child behavior are similar to those between parental attitudes and parent ratings of child behavior. Thus, the pattern of parental influence seems about the same whether children are observed in school or at home, even though teachers and parents view children from somewhat different vantage points. In retrospective light of these results, the low correlations between parent and teacher ratings do not seem as restrictive as they did when the analyses were begun. At the operational level, the results provide some justification for pooling such judgments when future analyses are planned, and, at the theoretical level, for conceiving once more of general child problem tendencies pervading behavior in many different kinds of situations.

Correlations for the Clinic families must be interpreted much more cautiously than those for the Kindergarten group. For a sample of 24 cases, a correlation must exceed .42 to achieve significance at the .05 level, and correlations of this magnitude are necessarily rare in the study of parent-child relationships. The data were also vulnerable to contamination from

knowledge of child problem status and are presented more in the interest of completeness and integrity than because of any new truths they reveal. Most of the "significant" correlations are between parental strictness and responsibility and the *absence* of personality problems among the children, possibly representing a relaxation of pressure by parents who perceive shyness, withdrawal, and other personality problems in their offspring. No other relations were confirmed, and most conclusions drawn from the study must be justified through reference to the comparison between Clinic and Kindergarten families and to the relatively "pure" correlations established within the larger Kindergarten sample.

Discussion

Let us now restate the hypotheses which guided this research and consider the results in the light of those propositions.

Hypothesis 1. The attitudes of fathers are at least as intimately related as the attitudes of mothers to maladjustive tendencies among children.

The prediction implicit in this hypothesis was clearly borne out. Of the 65 relations examined for fathers, 14 were "significant" at the .05 level. Among the same number of relations examined for mothers, only eight were "significant." One should not make too much of that disparity because of the multiple interdependencies among the variables, but the direction of the difference, at least, is interesting. The paternal influence, furthermore, appears just as strongly predominant for the present Kindergarten group as it has for previous samples (1, 5). This is important not only because data from the Kindergarten group are almost wholly free from contamination, but because the sexes are almost equally represented in the group. In fact, a few more kindergarten girls than boys participated in the study. Reservations arising from the preponderance of male children in groups studied before are thus mollified somewhat, and the critical importance of the father role is clearer than ever.

Hypotheses 2 and 3. Child personality problems are related to autocratic attitudes and lack of parental concern among fathers, while conduct problems are associated with maternal maladjustment and with permissiveness and disciplinary ineffectuality on the part of fathers.

The specificity which these propositions imply was not evident in the present results. Instead, a single attitudinal pattern, consisting most prominently of paternal harshness, was related indiscriminately to all three of the problem dimensions examined. Of many possible interpretations, the most tenable have to do with age differences between the present groups of children and the ones examined previously. The behavior of young children is massive, and it is stimulus-bound. Whatever differentiation of problem behavior five-year-olds display may be less a function of intrapersonal tendencies, determined by specific kinds of parental attitude, than of a general personality disturbance, shown now in one form and now in another, as

situational provocations and opportunities arise. From the study of older children, it seemed that paternal firmness was needed to prevent the occurrence of conduct problems in children. Limits are probably necessary all along, but, for these younger children, warmth, affection, and freedom from wrath seem more important in preventing problems of all kinds.

Summary

Previous research on parent-child relationships led to three guiding hypotheses for the present study.

1. The attitudes of fathers are at least as intimately related as the attitudes of mothers to maladjustive tendencies among children.

2. Personality problems among children are related to harsh, autocratic attitudes and lack of parental concern among fathers.

3. Child conduct problems are related to maternal maladjustment and to permissiveness and disciplinary ineffectuality on the part of fathers.

Fifty-three kindergarten children (Kindergarten group) and 24 children attending a guidance clinic (Clinic group) participated in the investigation. Their parents were interviewed and rated on five factors of parental attitude defined in previous research by Sears, Maccoby, and Levin (6). A problem checklist applicable to children was factor analyzed, yielding three dimensions of problem behavior. Independent ratings of these were obtained from teachers and parents. Attitudes for Kindergarten and Clinic parents were then compared, and correlations established between parental attitudes and child problems.

The first hypothesis stated above was clearly supported, and the importance of paternal attitudes strongly reaffirmed. The differentiation implied by the second two hypotheses, however, was not apparent in the results. Instead, a single parent attitude pattern was diffusely associated with personality problems, conduct problems, and autism as well. The most conspicuous elements in this pattern were the strict, cold, aggressive attitudes of the fathers. Among the older children examined in the previous research, a considerable amount of paternal firmness seemed to be necessary for the prevention or amelioration of child conduct problems. For the younger children of the present study, love and kindness seemed more generally important.

References

1. Becker, W. C., Peterson, D. R., Hellmer, L. A., Shoemaker, D. J., & Quay, H. C. Factors in parental behavior and personality as related to problem behavior in children. *J. consult. Psychol.*, 1959, 23, 107-118.

2. Eysenck, H. J. *The structure of human personality.* Wiley, 1953.

3. HIMMELWEIT, H. T. A factorial study of "Children's behavior problems." Unpublished ms., Univer. of London, 1952.

4. KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.

5. PETERSON, D. R., BECKER, W. C., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Parental attitudes and child adjustment. *Child Develpm.*, 1959, 30, 119-130.

6. SEARS, R. R., MACCOBY, E., & LEVIN, H. *Patterns of childrearing*. Row, Peterson, 1957.

DISCRIMINATION PERFORMANCE AS A BASELINE FOR INDIVIDUAL ANALYSIS OF YOUNG CHILDREN[1,2]

SIDNEY W. BIJOU*
University of Washington

It is well established that unequivocal relations between manipulated variables and instrumental or operant behavior may be demonstrated in single infrahuman organisms (4). The question raised here is whether such can be shown in research with young children. If reliable antecedent-consequent relations can be derived from individual analysis (the kind in which the subject serves as his own control), this type of design may be added to the investigator's repertory of skills with certain advantages. For example, individual analysis makes it possible to do considerably more research in situations in which only a limited number of subjects is available, it encourages functional analytic research with children with developmental anomolies, it makes possible controlled study of detailed and gradual developmental changes over long periods, and, because of its flexibility, it facilitates explorations into new procedures and functional relationships. These features and others have been discussed in some detail elsewhere (3).

Data from an individual preschool child will be presented consisting of baseline behavior and variations in performance following the introduction of a specific experimental condition. Two replications on two other young children will also be described.

The children were attending a private day-care center in Seattle and were observed in a mobile research laboratory parked adjacent to the school building.[3] The experimental room resembled a playroom and contained two tables with a toy on one and the apparatus on the other.

The toy, a standard commercial item (Playskool), consisted of a lever projecting from a wooden cage which contained a wooden ball. Striking

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] This paper with minor modifications was presented at a symposium, "Recent developments in experimental methods with children," at the annual meeting of the Amer. Psychol. Ass., Washington, D.C., September, 1958.

[2] This research was supported by grants from the National Science Foundation (G2141) and the National Institute of Mental Health (M-2208), U. S. Public Health Service.

[3] We wish to express our gratitude to Mrs. Charlotte Pearson and her staff of the Carrousel Nursery School for their exceptionally fine cooperation. We also wish to thank Robert Sharpley and Persis T. Sturges for assisting in the study.

the lever flipped the ball to the top of the cage. As it fell, it rattled to its original position and in a position to be tossed up again by the next response. The child could play with the toy at any time.

The apparatus was a colorfully painted box with two jewel lights, a metal handle protruding from the front panel, and a tray on one side for catching reinforcements falling from a chute extending through an opening in the wall separating the experimental and control rooms. Pressing the handle down (820 grams of pressure were required to make a complete response) was always followed by a relay click, and now and then by a reinforcer dispensed by a Gerbrands Universal Feeder in the control room. Responses are recorded on a Gerbrands cumulative recorder. A young lady (referred to as *A* in distinction to *E*, the experimenter) brought the child into the room, remained with him during the session, and returned him to school. She followed a routine for removing his coat, taking him to the center of the room, giving him instruction ("Here are some toys. You may play with them as long as you like."), and retiring behind a portable partition to read a book and carry out variables in procedures suggested by *E*. *E* was behind a one-way screen in the wall opposite to that where *A* was stationed. In addition to observing the child (*A* did not participate in observing the subjects), he monitored the apparatus and talked to *A* through a telephonic system which delivered signals audible only to *A*. Detailed descriptions of the mobile laboratory, equipment, and procedures have been published (1, 2).

Training was given to establish a discrimination marked by two minutes of rapid responding in the presence of an amber light on the left side of the panel and by two minutes of no responding to a blue light on the right side. When the amber light was on, reinforcements (trinkets, dime-sized cookies, sugar corn pops, and M & M candy, mixed in respective proportions of 5:3:3:1) were distributed on a variable ratio ranging from 10 to 90 with a mean of 50. Sessions were 20 minutes and intervals between sessions six days. Because of absences, the interval was extended to seven to eight days at times. After two sessions of stable performance, a toggle switch was installed over each light to enable the subject to turn out the lights.[4] Changes in discriminative behavior as a function of this alteration were observed.

Data from the first child, JDA, a boy, 4 years, 3 months of age, are shown in Figure 1 and the upper part of Table 1.

Chart A is a record of his seventh day and the second session in which stable discrimination was shown. Here, as well as in the other charts, the lower horizontal line is a time line which indicates occurrence of certain events, including change in the discriminative stimuli. When the line is down, the amber light is on for two minutes, and, when it is up, the amber light is off and the blue light is on for two minutes. The graph is a record of responses. When the apparatus handle is pressed all the way down, the

[4] To give the child a brief opportunity to explore the new attachments, working the switches during the first two minutes did not turn off the lights.

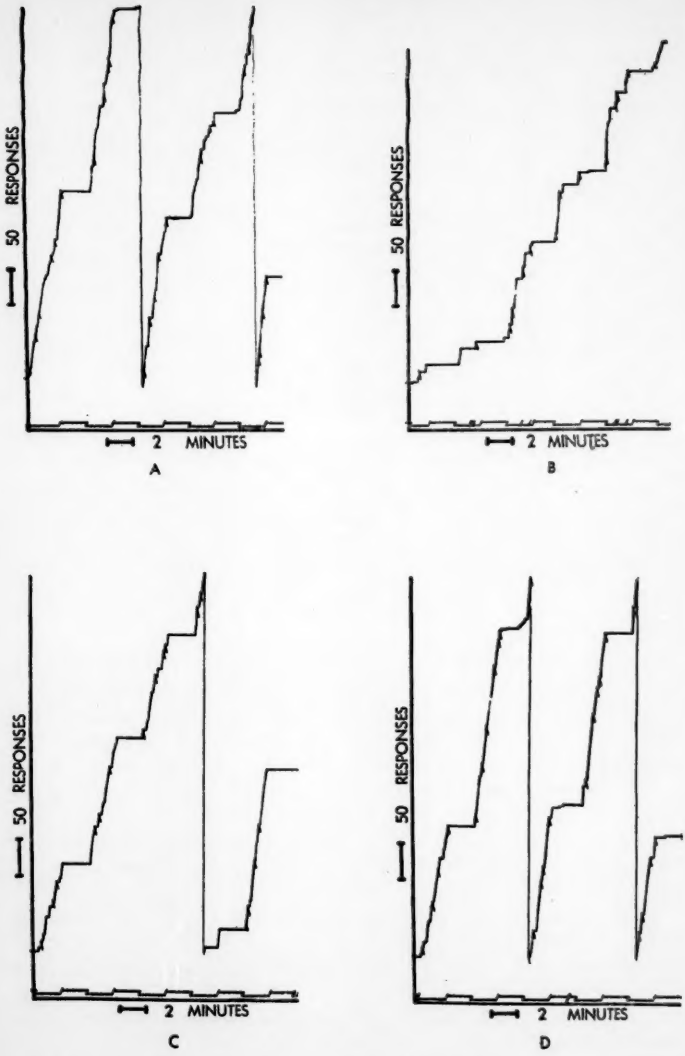

FIGURE 1—Performance of a 4-year, 3-month-old boy (JDA) on the baseline day (chart A) and three succeeding days (charts B, C, and D) after introduction of a new experimental condition.

TABLE 1

EXPERIMENTAL CONDITIONS, NUMBER OF RESPONSES, AND REINFORCEMENT SHOWN IN FIGURES 1, 2, AND 3

Session	*Days Since Previous Session*	*Experimental Condition*	*No. of Responses*	*No. of Reinforcements*
JDA, Boy 4 yrs., 3 mos.				
7	7	Standard	1260	25
8	6	Switches installed	581	13
9	6	Switches installed	852	16
10	8	Switches installed	1265	25
KBB, Girl 3 yrs., 10 mos.				
9	6	Standard	1590	32
10	6	Switches installed	1231	24
11	6	Switches installed	936	19
12	6	Switches installed	1094	21
PMC, Girl 4 yrs., 10 mos.				
9	6	Standard	1633	25
10	6	Switches installed	578	20
11	6	Switches installed	654	20
12	7	Switches installed	730	14

pen moves one-hundredth of an inch upward from the time line; when the handle is inactive, the pen produces a line parallel to the time line. The result is a cumulative curve. Pauses are readily perceptible and rate of responding is shown by the slope—the faster the rate, the steeper the slope. Five hundred and fifty responses are required to move the pen across the paper. When the pen reaches the upper terminus, it automatically glides back to the starting position. Resets are indicated by thin lines parallel to the ordinate. One other feature of the response curve: each reinforcement is indicated by a pip drawn roughly perpendicular to the slope of the curve, like a pip on the time line.

Chart A shows that during the period of amber light JDA maintains a steady rate of responding and during the period of blue light he stops. It may be noted that at the end of the second, third, and fourth blue light periods he made a few responses suggesting that he had discriminated the two-minute period. He made a total of 1260 responses and received 25 reinforcements.

Prior to session 8, the toggle switches were installed. JDA was brought into the room in the usual manner and given the usual instruction: "You may play with the toys as long as you like." As shown in chart B, his per-

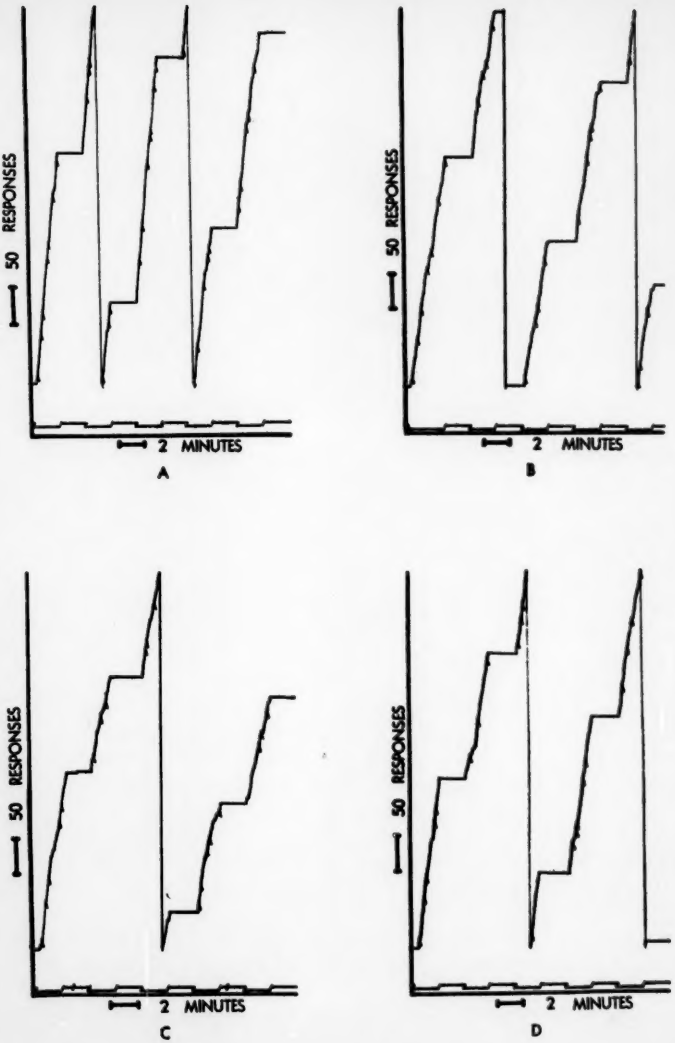

FIGURE 2—Performance of a 3-year, 10-month-old girl (KBB) on the baseline day (chart A) and three succeeding days (charts B, C, and D) after introduction of a new experimental condition.

formance was much poorer than in session 7. (Pips on the time line indicate the time and number of relay clicks made by *E* to call *S*'s attention to the amber light.) One index of the difference between session 7 and 8 is the decrease from 1260 to 581 responses. Several times during the session JDA went to *A* and asked about the meaning of the switches. In conformance with procedures, she merely reflected his questions.

In session 9, chart C, JDA's performance was more efficient than in session 8. He made 852 responses and received 16 reinforcements.

In session 10, chart D, his performance was similar to that in session 7, the baseline day. (The pip during the third elevation of the time line indicates that the blue light was turned off.) He received the same number of reinforcements received in session 7, and he came within five responses of giving the same number of responses. The introduction of the switches showed greatest influence in the eighth and ninth sessions.

Data on the second child, KBB, a 3-year, 10-month-old girl, are shown in Figure 2 and the middle section of Table 1.

Chart A shows her performance on the ninth experimental day, the second day of stable discrimination. She made 1590 responses and received 32 reinforcements, a performance which approaches maximum efficiency.

Before session 10, the toggle switches were added. As shown in chart B, the child's rate of responding in the presence of the amber light was decreased. Number of responses dropped from 1590 to 1231 and the number of reinforcements from 32 to 24.

In session 11 her performance was inferior to that in 10 as shown in chart C. Responses fell from 1231 to 936 and reinforcers from 24 to 19. But in session 12, chart D, some improvement was shown with 1094 responses and 21 reinforcers.

Findings on the third child, PMC, a girl 4 years, 10 months of age, are given in Figure 3 and the lower section of Table 1. The record of the ninth experimental day is shown in chart A. Like KBB, it took nine sessions for her to reach stable discriminative behavior. She made 1633 responses and received 25 reinforcements.

Before the tenth session, the switches were installed, and, as shown in chart B, her behavior in the tenth period was less precise than in the ninth. (The pip in the second depression on the time line indicates time-out to go to the bathroom. The one in the third elevation period indicates that the blue light was turned off.) Due to a conflict with the school schedule, the session had to be terminated at the end of 18 minutes. In comparison with the first 18 minutes of session 9, number of responses emitted fell from 1621 to 578 and number of reinforcements received from 25 to 20. There was some improvement in session 11 as shown in chart C. Variability in the initial phase of performance is correlated with turning off the amber light on the first and second presentations. In session 12, chart D, her responding continued to be uneven as compared with session 9, chart A. (She turned the amber light off on the first and second appearances and the blue off on the

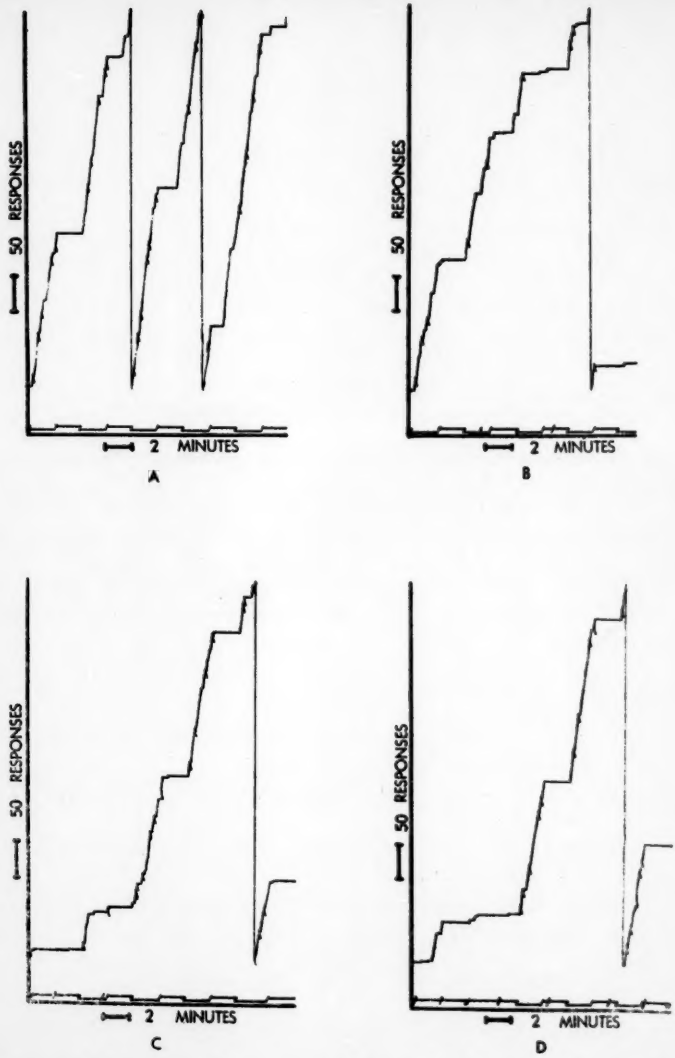

FIGURE 3—Performance of a 4-year, 10-month-old girl (PMC) on the baseline day (chart A) and three succeeding days (charts B, C, and D) after introduction of a new experimental condition.

second, third, and fourth presentations.) She made 730 responses and received 14 reinforcements.

Summary

The question raised here is whether unequivocal relations can be shown in the *individual* performance of young children. Data from three preschool children were presented to show that stable performance of two minutes of responding in the presence of an amber light and two minutes of no responding in the presence of a blue (a pattern which took from six to eight days to establish) was altered by the introduction of a pair of toggle switches allowing the child to control the lights. The discriminative performance of one child (JDA) was most affected the first day after installation of the switches, less on the second, and by the third day it was similar to the baseline performance. The behavior of the second child (KBB) showed the opposite trend. Alteration in discriminative behavior after experimental modification was slight on the first day and increased on the second and third days. Recovery began on the fourth day. The third child (PMC), like the first, showed the greatest change on the first experimental day. However, her "recovery rate" was slower than the first child's for, in the fourth session after installation of the switches, her performance was still far below the efficiency shown on her baseline day. These data are taken as support for the contention that individual experimental analysis is feasible with young children since a clear functional relationship has been shown between a stable baseline performance and the introduction of a special stimulus condition.

References

1. Bijou, S. W. Methodology for an experimental analysis of child behavior. *Psychol. Rep.*, 1957, 3, 243-250.
2. Bijou, S. W. A child study laboratory on wheels. *Child Develpm.*, 1958, 29, 425-427.
3. Skinner, B. F. A case history in scientific method. *Amer. Psychologist*, 1956, 11, 221-233.
4. Skinner, B. F. The experimental analysis of behavior. *Amer. Scientist*, 1957, 45, 343-371.

CULTURAL STEREOTYPES AND THE VALIDITY OF INTERVIEWS FOR RESEARCH IN CHILD DEVELOPMENT[1]

JOAN McCORD *and* WILLIAM McCORD*
Stanford University

One of the fundamental problems in the study of child development is the difficulty in obtaining valid information regarding childrearing practices. Most studies of child development are based upon interviews with the parents. Yet the interview as a technique for analyzing parent-child or mother-father relations has been subjected to serious criticism. After a comparison of the results of two pieces of similar research on social class differences in approaches to child training which yielded different results, Havighurst and Davis were forced to conclude: ". . . this reinforces the conviction of the present writers that the interview is not nearly so good as participant observation for securing data both on the behavior and the attitudes of parents toward their children" (1, p. 442). In their recent, extensive study of child development, Sears, Maccoby, and Levin (despite the care with which the interviews were conducted) had to remind their readers of the possibility that some of their findings "may have resulted from a response set created by the interview itself" (6, p. 63). And, in an article summarizing research on interview techniques, Eleanor and Nathan Maccoby have pointed out: "Validity studies all suggest that when people are interviewed directly concerning behavior about which there is a strong expectation of social approval or disapproval, and in which there is considerable ego-involvement, respondents tend to err in the direction of idealizing their behavior" (2, p. 482). Although researchers agree on the possible dangers inherent in interviewing, there has been little research directed toward the identification of the specific areas in which considerable ego-involvement can be expected.

* Sociology Department, Stanford University, Stanford, California.

[1] The authors wish to express their gratitude to the Ella Lyman Cabot Foundation, the Harvard Laboratory of Social Relations, and the National Institute of Mental Health (Grant M-2647) for their generous financial support of this research.

Child Develpm., 1961, 32, 171-185.

Even though most researchers are agreed that important biases may mar the results of studies based on interviews, practical alternatives are by no means easy to find. In this article, we will attempt to suggest those areas of childrearing in which the interview may be an adequate technique, as well as indicating certain systematic interview biases which are apparently due to cultural values or to strong expectations of social approval or disapproval.

As many researchers have commented, for an ideal study of childrearing, one would like to place an invisible, unbiased observer in the family for a period of several years. The Cambridge-Somerville Youth Study provided what is probably the closest approximation to such an invisible, unbiased observer. Begun in 1935, under the guidance of Richard Clark Cabot, this project was designed primarily for the purpose of preventing delinquency and of testing the effectiveness of a preventive program. Although the attempt to deter delinquency seems to have failed, the project resulted in a unique fund of information regarding childhood behavior, family interaction, and the transmission of values. Five hundred ten lower- and lower-middle-class boys were included in the study. These boys were divided into a "treatment" and a "control" group after data had been obtained from medical and psychological examinations and reports from home visitors who interviewed the mother regarding her childrearing practices. On the basis of this information, each boy was matched to another as closely similar in background and in personality as was possible. One in each pair was then selected, by toss of a coin, to receive the services of the Youth Study. The treatment program began in 1939 and continued until 1945. The "control boys" and their families were interviewed periodically during these years. The division between treatment and control groups and the attendant difference in available information about childrearing practices provided the basis for the present study. Those boys and their families who received treatment (hereafter called the "observation group") were directly observed by several workers over a period of more than five years. The control group (hereafter called the "interview group") were interviewed several times in the standard fashion. Thus, the information which emerged from this study was derived from fundamentally different sources.

Method

The Observation Group

For the boys who received treatment, complete records were reported after each contact by the counselor. Counselors gathered information from teachers, ministers, parents, and neighbors detailing the backgrounds of each of their boys. More importantly, the counselors repeatedly visited the boys and their families. The families were visited usually weekly or bi-monthly, and visits continued for a median of five and three-quarters years. The counselor-observers (except for one who was a nurse) were trained social workers, psychologists, or youth workers. They described in detail

whatever activities they observed in the home, at school, during chance meetings on the streets, and in the Cambridge-Somerville Youth Study center. They reported conversations with parents, friends, neighbors, and, of course, the subjects themselves.

In 1957, these raw observations were coded by trained researchers. Rigid definitions of the categories and repeated tests of interrater agreement resulted in a relatively high degree of reliability and, presumably, the elimination of "the halo effect" in rating.

The Interview Group

For the control boys, records had been kept of the original interviews with the mothers which had been used for matching cases. These were semi-structured. Although the exact order and specific wording of the questions were left unspecified, the interviewers were to fill out information sheets including:

> A. History of subject—eating habits (What good eating habits does subject have? What poor habits?), sleeping habits (How long does he sleep? Do you have trouble getting him to bed?), recreational habits (What does subject like to do?), attitude toward school (Do you have much trouble with subject about truanting? What does he like best about school? What does he like least?), attitude toward discipline (What does subject do when you have to punish him?), attitude toward parents (How does subject feel about his father? Give an example. How is he with you?), personality (What kind of boy is subject? How do the neighbors get along with him?).
>
> B. Father's history—occupational background, education, personality (What is your husband like? Do you have trouble getting him to do things you want?), methods of discipline (How does he get subject to do what he should? What happens if subject is bad? Can you give an example?), attitude toward subject (How does your husband get along with subject? Does he have plans for subject's future? Do they play together often?)
>
> C. Mother's history—same as for father.

In addition, at least one interview with the boy and three unstructured interviews with the mother had been gathered for each case. These latter pertained primarily to questions of disciplining the child, the mother's attitude toward the subject and his siblings (if any), and a general picture of family life.

The interview records, too, were categorized with high interrater agreement.

Since the same individuals gathered the records on the observation group and the interview group, it seems reasonable to assume that, if their attitudes affected the nature of the data recorded, these errors would be similar for the two groups.

These two types of records—those gleaned from continuous reports of direct observation over a number of years and those summarizing several intermittent interviews with the mother—furnish the data for the present analysis.

We reasoned that if the observation records were valid and if the interviewed families were actually similar to the observed families in behavior and attitudes, any apparent differences in reported information might be correctly attributed to systematic interview biases. Therefore, before turning to the results of this comparison, we must consider two questions: (a) evidence for the validity of the information gathered about the observation group, and (b) evidence that the childrearing practices of the interview group were in fact similar to those in the observation group.

The Problem of Validity

The face validity of the observation records is considerable. The families had been visited repeatedly by trained observers, the observations had been recorded almost immediately after each visit, and the families were not aware of the fact that records were being kept. (As far as the parents were concerned, the Cambridge-Somerville Youth Study was a "service" organization.)

Counselor turnover, considered a potential disadvantage for treatment, proved an advantage for research: the vast majority of subjects were observed by two to four counselors. Thus, possible individual observer biases tended to cancel one another.

Perhaps the most important evidence for the validity of the observation records is the highly significant relations yielded by these records between childrearing practices and the criminality, alcoholism, and psychoses of the boys *as adults* (3, 4, 5). Thus, we know that the observation records have good "predictive power."

TABLE 1

ECOLOGICAL BACKGROUND OF SUBJECTS

	Observation Group ($N = 255$)	*Interview Group* ($N = 255$)
	PERCENTAGE	
Race		
White	91	91
Negro	9	8
Mixed	0	1
Father's Birthplace		
United States	49	44
Foreign-born	49	52
Not ascertained	2	4
Neighborhood in Childhood		
Residential	37	35
Transitional	62	62
Not ascertained	1	3

TABLE 2

CULTURAL BACKGROUND AND AGE OF PARENTS

	Observation Group ($N = 255$)	*Interview Group* ($N = 255$)
	PERCENTAGE	
MOTHER		
Religion		
Catholic	62	57
Protestant	36	38
Other	0	2
Not ascertained	2	3
Education		
College graduate	1	0
High school graduate	10	8
Some high school	15	13
Grammar school graduate	17	20
Some grammar school	33	32
No formal education	11	9
Not ascertained	13	18
Age of Parent When Boy Was Born		
Under 20	8	6
20 - 24	21	19
25 - 29	21	22
30 - 34	25	19
35 - 39	13	14
40 - 44	4	7
Over 44	1	1
Not ascertained	7	12
FATHER		
Religion		
Catholic	61	54
Protestant	30	34
Other	2	2
Not ascertained	7	10
Education		
College graduate	1	1
High school graduate	11	9
Some high school	12	12
Grammar school graduate	19	19
Some grammar school	31	29
No formal education	13	9
Not ascertained	13	21
Age of Parent When Boy Was Born		
Under 20	0	0
20 - 24	11	11
25 - 29	23	20
30 - 34	24	22
35 - 39	17	18
40 - 44	11	13
Over 44	7	5
Not ascertained	7	11

The Question of Comparability

Three types of evidence indicate that the observation group and the interview group were very similar in attitudes and behavior.

First, the two groups gave similar responses when interviewed at the beginning of the project. In fact, the two groups were matched largely on the basis of information given in these interviews. Although systematic biases tend to interfere with the validity of information gathered by interviews and are probably related to actual childrearing techniques, the fact that the two groups responded in a similar fashion indicates that similar processes controlled the nature of these responses.

Secondly, we compared the two groups using "objective" variables which were least likely to be subject to bias and yet might be correlated with differences in childrearing behavior. The figures are presented in the following charts.

We found no significant differences[2] between the two groups when we compared their racial background, their father's birthplace, or the neighborhoods in which the boys had been reared (*see* Table 1).

We found practically no differences between the groups in the cultural background or age of their parents (*see* Table 2).

The similarities between the groups regarding education of the father and place of residence point to equal social status between the groups. Nevertheless, we found a slight, but statistically nonsignificant ($\chi^2 = 2.80$), difference in the occupational status of the boys' fathers. This difference, we believe, might be attributable to misleading statements of the interview group since it is in a "culturally approved" direction (*see* Table 3).

The striking similarities in family background found when we compared race, neighborhood, father's birthplace, religion, education, and parental age lead one to believe that the two groups closely resembled each other in terms of objective social characteristics.

We considered the possibility that the number of children in the family, the ordinal position of the child, and the sex of his siblings might affect childrearing practices. Again, we found that the observation and interview groups had essentially similar compositions (*see* Table 4).

Thus, by relatively objective criteria, the observation group and the interview group were basically similar; one would presume, therefore, that the two groups would also be similar in childrearing techniques.

Our third approach to ascertaining the comparability of the two groups was based upon the relationships we had found between childrearing practices and adult social deviance (3, 4, 5). We reasoned that, if adult social deviance is strongly related to the behavioral observations made in childhood, then a similarity in adult deviance patterns should indicate a similarity in childrearing techniques. The incidence of adult deviance of the boys (as

[2] Tests of significance were two-tailed, using $p < .05$ as the minimum standard for asserting significance.

TABLE 3

FATHER'S OCCUPATION

	Observation Group ($N = 255$)	*Interview Group* ($N = 255$)
	PERCENTAGE	
Professional	1	1
White collar	8	12
Skilled tradesman	29	29
Unskilled worker	62	58

TABLE 4

FAMILY COMPOSITION OF SUBJECTS

	Observation Group ($N = 255$)	*Interview Group* ($N = 255$)
	PERCENTAGE	
Number of Children in Family		
One	5	5
Two	13	12
Three	15	14
Four	15	16
More than four	52	53
*Ordinal Position of Boy**		
Only child	5	6
Twin	2	2
Oldest	17	15
Middle	48	49
Youngest	28	28
Sex of Siblings		
Male only	14	12
Female only	13	13
Male and female	66	67
None or twin	7	8

* A boy was rated as an only child if no other siblings were living in the home.

TABLE 5

RECORD OF SOCIAL DEVIANCE IN ADULTHOOD*

	Observation Group	*Interview Group*
	NUMBER OF SUBJECTS	
Noncriminal alcoholics	11	10
Nonalcoholic criminals	44	44
Alcoholic criminals	18	12
	73	66

* Subjects were considered alcoholics if they had been members of A.A., received medical treatment for alcoholism, or had been convicted more than once for public drunkenness; they were considered criminals if they had been convicted by the courts for theft, assault, or sex crime at least once.

ascertained through perusal of community records when the boys reached adulthood) was almost identical for the two groups (*see* Table 5).

In view of the striking similarities between the two groups on all of these measures, we are reasonably confident that any major differences in reports on childrearing practices between the two groups as reported through interviews resulted from biases inherent in the interviews, rather than actual differences. On the other hand, for those variables on which we found little variation between the reports of the participant observers and the interviewers, there is indirect evidence that the interviews produced valid information.

RESULTS

In analyzing the similarities and differences, we will consider three phases of family life which affect childrearing: first, the emotional tone of the family; second, leadership and the division of responsibility between the parents; and third, techniques used for controlling the child.

The Emotional Tone of the Family

For the observation group, of course, actual behavior was observed over an extended period of time. For the interview group, ratings were based on summary judgments of a minimum of four interviews. Measures were designed to record the emotional relation between the mother and child, the father and child, and between the father and mother. *Some, but not all, of these measures seemed to be subject to interview bias.*

When we looked at the comparative ratings of the parents' attitudes toward their sons, we found that for both mothers and fathers *significantly less active rejection was revealed through the interviews* (*see* Table 6).

TABLE 6

PARENTAL ATTITUDES TOWARD THEIR SONS*

	MOTHER		FATHER	
	Observation Group ($N = 250$)	*Interview Group* ($N = 241$)	*Observation Group* ($N = 216$)	*Interview Group* ($N = 210$)
	PERCENTAGE		PERCENTAGE	
Warm	50	46	30	32
Passive	32	44	46	55
Rejecting	18	10	24	13
	$\chi^2 = 8.1$; $p < .005$		$\chi^2 = 8.0$; $p < .005$	

* "Warm" parents evidenced active affection for their children through open approval and pleasure in being with them; "passive" parents had little interaction with their sons although they may have been concerned for their welfare; "rejecting" parents openly demonstrated their dislike for their offspring. Agreement of two independent raters on a random sample of the observation group was 83 per cent for the mothers' and 77 per cent for the fathers' attitudes; on a random sample of the interview group, agreement was 84 per cent for the mothers' and 76 per cent for the fathers' attitudes.

Apparently, even repeated interviews failed to differentiate passive parents from those who actively rejected their children. Yet there seems to have been little systematic error in distinguishing the actively affectionate parents from those who were *either* passive or rejecting.[3]

Not surprisingly, we found that *the interviews indicated significantly fewer negative attitudes of the boys toward their parents* (*see* Table 7). Sixteen per cent of the total 255 observed subjects had revealed disapproving or fearful attitudes toward their mothers and 33 per cent had evidenced dislike of their fathers. In the interview group, however, only 7 per cent gave evidence of negative attitudes toward their mothers and 19 per cent toward their fathers.

Despite the inadequacies of the interview in detecting rejection between parent and child, there do seem to be indices of the family's emotional tone which are relatively free from systematic bias. There was very little difference between the observation and the interview groups either on the measure of parental attitudes toward each other or on the amount of conflict between the parents. Approximately one-fourth of each group of parents were rated as antagonistic and approximately one-third evidenced intense conflict (*see* Table 8).

Thus, in terms of emotional interaction, the interview appears to be more reliable as a technique for judging the attitudes of parents toward each

[3] Hence, we would hazard the suggestion that scales of attitudes of parents toward their children be based on positive signs of warmth rather than upon signs of rejection.

TABLE 7

BOYS' ATTITUDES TOWARD THEIR PARENTS*

	Attitude toward Mother		Attitude toward Father	
	Observation Group ($N = 220$)	*Interview Group* ($N = 208$)	*Observation Group* ($N = 203$)	*Interview Group* ($N = 187$)
	PERCENTAGE		PERCENTAGE	
Favorable	82	91	59	74
Unfavorable	18	9	41	26
	$\chi^2 = 7.4$; $p < .01$		$\chi^2 = 10.7$; $p < .005$	

* The boy's attitude toward his parent was considered "favorable" if there was some evidence of general approval and "unfavorable" if the boy appeared to fear, disapprove, or disdain his parent. Interrater agreement for the observation group was 93 per cent for attitude toward mother and 80 per cent toward father. For the interview group, interrater agreement was 88 per cent toward mother and 96 per cent toward father.

TABLE 8

PARENTAL WARMTH AND CONFLICT*

Relation between Parents			Degree of Parental Conflict		
	Observation Group ($N = 205$)	*Interview Group* ($N = 182$)		*Observation Group* ($N = 214$)	*Interview Group* ($N = 207$)
	PERCENTAGE			PERCENTAGE	
Affectionate	60	62	Little	30	34
Indifferent	12	9	Some	33	35
Antagonistic	28	29	Considerable	37	31

* Parents were considered "affectionate" if there was evidence of mutual concern for one another's welfare, "indifferent" if the parents seemed unconcerned about each other, and "antagonistic" if they seemed to take pleasure in discomfitting one another. Interrater agreement was 73 per cent for the observation group and 80 per cent for the interview group. Interrater agreement regarding amount of parental conflict was 73 per cent in the observation group and 68 per cent in the interview group.

other than of attitudes between parents and their children. These findings suggest that the admission of parental conflict and antagonism is more "permissible" than is a confession of parent-child antagonisms.

Leadership and Division of Responsibility

Who makes the decisions in the family? Who carries them out? The answers to these questions impinge upon the child's perception of the world and, to a considerable extent, have been found to be related to his adult

TABLE 9

ROLE DIFFERENTIATION*

	Observation Group ($N = 203$)	*Interview Group* ($N = 205$)
	PERCENTAGE	
Highly differentiated	49	60
Not highly differentiated	51	40
	$\chi^2 = 5.7$; $p < .02$	

* Interrater agreement was 70 per cent for the observation group and 92 per cent for the interview group.

behavior (3, 4, 5). In assessing the answers to these questions, we found several apparent interviewing biases.

One aspect of the division of responsibility pertains to role differentiation. Do mother and father share in such activities as caring for the children, earning a livelihood, preparing meals, cleaning house? Considering only those families in which both mother and father were available as models, we found that *the interview reports indicated a greater degree of differentiation than was found among the observation group* (*see* Table 9).

Parental dominance was a measure designed to show which parent made important decisions for the family (such as where to go on vacations, how to spend money, what to do when crises arose). Judgments based on interviews differed significantly from those based on direct observation: *The interviewed mothers tended to portray their husbands as more dominant* (*see* Table 10). These differences between the interview and the observation

TABLE 10

PARENTAL DOMINANCE*

	Observation Group ($N = 254$)	*Interview Group* ($N = 228$)
	PERCENTAGE	
Father	39	49
Mother	32	25
	$\chi^2 = 5.2$; $p < .025$	
Equal	7	7
Only one parent at home	22	19

* Interrater agreement was 83 per cent for the observation group and 84 per cent for the interview group.

TABLE 11

PARENTAL ESTEEM*

	Esteem of Father for Mother		Esteem of Mother for Father	
	Observation Group ($N = 186$)	*Interview Group* ($N = 165$)	*Observation Group* ($N = 223$)	*Interview Group* ($N = 199$)
	PERCENTAGE		PERCENTAGE	
Moderate or high	63	71	54	66
Low	37	29	46	34
	$\chi^2 = 2.7$; *ns*		$\chi^2 = 6.9$; $p < .01$	

* Interrater agreement of esteem of father for mother was 80 per cent for the observation group and 96 per cent for the interview group. Agreement between raters on esteem of mother for father was 90 per cent on the observation group and 92 per cent on the interview group.

groups illustrate the American emphasis on differentiating the role of husband and wife, with the father occupying the role of family leader.

Further indication of the tendency to portray the father in the role of leader was found in the analysis of reports on the amount of respect the parents had for one another (*see* Table 11). One would expect that, if interview bias was due primarily to a direct attempt to give oneself status, there would be considerable discrepancy between interviews and observations in the measure of the father's esteem for the mother—i.e., one would expect the mothers to indicate that their husbands held *them* in high esteem. Although there was some tendency in this direction, the difference between the interview and the observation records was not statistically significant. On the other hand, there did appear to be some interviewing bias in measuring the mother's esteem for the father. Thus, although there was a slight tendency for the mothers to overvalue themselves, there was a much stronger tendency for them to indicate higher esteem of their husbands—in other words, their interviews presumably reflected a cultural stereotype of the wife "respecting" her husband.

The same phenomenon occurred in the comparative analysis of the "leadership roles" of the parents. For this measurement, each parent was considered individually (e.g., both parents might be "leaders" although one was dominant); the interviewed mothers more often pictured their husbands as "dictators" or "leaders" (*see* Table 12). It should be remembered that judgments for the observation group were based on reports of the actual interaction among family members and that judgments for the interview group were based on at least four interviews with the mother and not simply on their responses to one question. Consciously or unconsciously, therefore, the interviewed mothers must have wanted to create the general impression

TABLE 12

PARENTAL ROLES*

	MOTHER		FATHER	
	Observation Group ($N = 247$)	*Interview Group* ($N = 231$)	*Observation Group* ($N = 213$)	*Interview Group* ($N = 203$)
	PERCENTAGE		PERCENTAGE	
Dictator	4	5	14	17
Leader	65	63	45	54
Passive	15	15	36	24
	$\chi^2 = 0.7$; *ns*†		$\chi^2 = 7.3$; $p < .05$†	
Ambiguous	16	17	5	5

* Parents were considered "dictators" if they made all or most decisions without consulting one another and with almost complete authority. They were considered "leaders" if they accepted responsibility for many decisions. They were considered "passive" if they made few decisions; and their role was considered "ambiguous" if they were rarely involved in family affairs (e.g., if they came home only periodically) or if they made decisions and yet refused to take responsibility for them. Interrater agreement on mother's role was 73 per cent for the observation group and 76 per cent for the interview group; for father's role, interrater agreement was 73 per cent for the observation group and 64 per cent for the interview group.

† $df = 2$.

that their husbands were leaders in the home.[4] This, we can assume, reflects a generalized ideal for lower- and lower-middle-class American families.

Techniques of Control

There are, of course, many facets to the technique of control by which the parent transmits values to the child. There was no difference between the two groups in reports on degree of supervision exercised by the parents, in their reports as to which parent took major responsibility for discipline, or in judgments concerning the mother's method of discipline. The interviewed mothers did, however, tend to report their husbands as less punitive (*see* Table 13).

The greater bias found regarding the father's technique is open to several interpretations. Two seem particularly plausible. First, the results might be further evidence of the tendency of the mothers to "upgrade" their husbands. Second, the results might indicate a stereotype of ideal parenthood in which the mother, "because she must," is permitted to utilize physical punishments, but the father is expected to control his children through the authority of his position as "head of the home."

[4] Apparently there is less cultural stereotype regarding discipline. In both groups, approximately 60 per cent of the mothers were regarded as the major disciplinarian.

TABLE 13

TECHNIQUE OF DISCIPLINE*

	MOTHER		FATHER	
	Observation Group ($N = 247$)	*Interview Group* ($N = 218$)	*Observation Group* ($N = 206$)	*Interview Group* ($N = 197$)
	PERCENTAGE		PERCENTAGE	
Punitive	49	45	62	52
Nonpunitive	51	55	38	48
	$\chi^2 = 0.1$; *ns*		$\chi^2 = 4.4$; $p < .05$	

* Severe spankings and harsh verbal threats were considered as "punitive." Interrater agreement for mother's discipline was 97 per cent on the observation group and 92 per cent on the interview group; for father's discipline, interrater agreement was 77 per cent on the observation group and 84 per cent on the interview group.

SUMMARY

In an attempt to assess the validity of interviews in the investigation of child development, two groups of 255 families were compared. Information about one of these groups was gathered through extensive observation over an average of five and three-fourths years; information about the other group was obtained through a series of interviews with the mother in the home. Originally, the two sets of families had been matched to insure their comparability; analysis revealed that the "observation group" and the "interview group" were essentially similar in race, birthplace of father, neighborhood, religion, education of parents, age of parents, number of children in family, ordinal position of children, and sex of children. Thus, in terms of "objective" criteria, there were no major differences between the two sets of families.

We assumed, therefore, that any contrasts in reports on childrearing techniques resulted from interview biases. Analysis revealed that the interview group more often reported that: the mother did not actively reject her son ($p < .005$); the father did not actively reject his son ($p < .005$); the boy had a favorable attitude toward his mother ($p < .01$); the boy had a favorable attitude toward his father ($p < .005$); roles within the family were clearly differentiated ($p < .02$); the father was the dominant person in the family ($p < .025$); the mother regarded the father with high esteem ($p < .01$); the father played the role of "leader" for the family ($p < .05$); the father was nonpunitive in disciplining the son ($p < .05$).

On the other hand, there were no major differences in reports on such variables as: the parents' affection for each other, the parents' conflict with one another, the esteem of the father for the mother, the degree of super-

vision exerted by the parents over the child, the mother's technique of discipline.

It would appear, therefore, that the validity of the interviews was marred by the parents' tendency to make their picture of family life conform to cultural stereotypes. Specifically, it seems that the interviewed parents viewed the "ideal" family as one in which the parents loved their children (and vice versa), the father directed family affairs, and the father was a "decent" person who treated his son with kindness. Apparently, the interviewed mothers were more willing to place themselves in an honest, if unflattering, light than they were to undermine the "respectability" of their husbands. Thus, for a variety of possible reasons—perhaps because prolonged contact between the parents and the observers led to a greater willingness for the parents to reveal themselves, perhaps because the parents found it impossible to keep up a facade throughout the five-year period—it would seem that the information derived from direct observation of the families offered a more valid picture of their actual practices.

It is possible that changes in parent-child relations over a five-year period might account for some of the differences. One might argue, for example, that it is more likely for parents to reject their children as boys enter the conflictful adolescent period than it is to reject them when they are small children. By this argument, differences between the interviewed and observed groups would not be due to biased reporting, but would be an outcome of the maturational process. Although this interpretation could fit some of the findings, it would not explain others, such as the tendency, within the observed group, for less role differentiation in the family.

References

1. Havighurst, R. J., & Davis, A. A comparison of the Chicago and Harvard studies of social class differences in child rearing. *Amer. sociol. Rev.*, 1955, 20, 438-442.
2. Maccoby, E. E., & Maccoby, N. The interview: a tool of social science. In G. Lindzey (Ed.), *Handbook of social psychology*. Addison-Wesley, 1954. Pp. 449-487.
3. McCord, W., & McCord, J. (with Zola, I. K.) *Origins of crime*. Columbia Univer. Press, 1959.
4. McCord, W. & J. (with J. Guderman) *Origins of alcoholism*. Stanford Univer. Press, 1960.
5. McCord, W., Porta, J., & McCord, J. The familial genesis of psychoses. *Psychiatry*, in press.
6. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing*. Row, Peterson, 1957.

EFFECTS OF MATERNAL EMPLOYMENT ON THE CHILD[1]

LOIS WLADIS HOFFMAN*
University of Michigan

Empirical studies of the effects of maternal employment have long suffered from a paucity of adequate controls. It was the lack of controls, for example, which led to the long-standing belief that maternal employment was strongly associated with juvenile delinquency. In fact, both maternal employment and delinquency were associated with social class and with broken homes; and, when the latter variables were controlled the relationship disappeared, suggesting it had been a spurious one (3).

Yet, the new studies on maternal employment, despite their more adequate controls, have often produced more confusion than clarification. Maternal employment has been associated with *less* delinquency, *more* delinquency, withdrawal behavior, dominance behavior, and nothing at all (1, 4, 9, 10, 11). Furthermore, there are existing theories to account for each of these findings. Thus, delinquent and hostile-assertive behavior might result from parental neglect and lack of sufficient discipline due to the mother's absence from the home; from hostility on the part of the child at being deserted; or, in conjunction with Henry's notion (5) that father discipline leads to the expression of hostility outward whereas mother discipline leads to intrapunitiveness, from father discipline replacing mother discipline.

Withdrawal and dependency responses by the child might also be expected. Loss of mother through death or hospitalization has been associated

* Research Center for Group Dynamics, University of Michigan, Ann Arbor.

[1] This paper was presented at the National Council on Family Relations, Ames, Iowa, 1959. The research was supported by the National Institute of Mental Health.

Child Develpm., 1961, 32, 187-197.

with overdependency—the young child perhaps seeking in this crisis a return of the maternal relationship he had in the past. Something similar might occur as the result of the mother's spending a great deal of time outside the home working. Still another reason for withdrawal and dependency might be that the working mother feels guilty about working outside the home and responds with a pattern of "smother love" or overprotection. Such a pattern could alleviate the mother's guilt. In addition, if it is true that working mothers have greater power needs than nonworking mothers, it would also be an appropriate pattern for expressing power over the child in a way that appears to be benign and admirable.

Some, such as Cyrus (2), would expect working mothers to be less frustrated and more capable of warm interaction. Others would say that the energy output required by trying to fill two jobs would make the mothers more irritable with their children.

Instead of speculating as to which of these theories is correct, the present paper will suggest and illustrate an approach in which the question is: *When does one pattern operate and when does another?*

Each theory assumes something about the working situation that is sometimes, but not always, correct; and, when certain assumptions hold, different mother-child interactions will result and the child will be affected accordingly. One important factor implicit in several of the theories involves the question of whether or not the mother enjoys her employment. It is this difference that the present study has selected for focus. It was expected that the effects of employment on the mother-child relationship, and consequently on the child, would be different when the working mother enjoyed work than when she did not.

Two of the theories mentioned above, the "guilt-overprotection" theory and the "neglect" theory, lead to very different predictions about the mother's behavior toward the child, and, although both theories predict that the working mother's child will be disturbed, the pattern of disturbance predicted by each is very different. Both these theories served as guides in the present investigation. The "guilt-overprotection" theory was expected to operate for the working women who enjoyed work, and the "neglect" theory for the working women who did not enjoy work.

Only the working woman who enjoys her employment should be guilty about it, and as such it was predicted that such a woman would try to compensate for her employment by showing a great deal of affection toward the child, disciplining the child adequately to leniently, and being careful that the child should not be inconvenienced by the fact that she works, e.g., by having to help with household tasks. The child correspondingly should be relatively nonhostile, nonassertive, and, if this parental pattern were carried to the point of overprotection, somewhat withdrawn and passive. The working woman who dislikes her work should, on the other hand, be quite guiltless. As such, she was expected to show more of a withdrawal from the maternal role—showing less positive affect and less disciplining. She should

feel freer to inconvenience the child—specifically to expect the child to help with household tasks. The child, feeling somewhat resentful and lacking sufficient discipline, was expected to be assertive and hostile. The predictions were, then, that both groups would show different patterns than when the mother was not employed, but that these differences would not be the same.

Procedure

The sample included 176 white, *intact* families with at least one child in the third through sixth grades of three elementary schools in Detroit. There were 88 working-mother families and 88 nonworking-mother families matched on occupation of father, sex of child, and ordinal position, including the important status of being an only child. Except for sex of child, all matched variables were selected because they were known to be related to maternal employment (7) and because it was believed they might be related to the dependent variables. Pairs of families were matched by sex of child so that it might be possible to examine the relationship between maternal employment and the dependent variables separately for boys and girls, to be certain that combining was legitimate. All statistical comparisons are between working and nonworking populations, and with a few exceptions which will be pointed out later, *all comparisons are for matched pairs.*

The data reported here are based on questionnaires filled out by the children, interviews with the mothers, teacher ratings, and a classroom sociometric. Each of these measures will be taken up in turn when the findings are presented.

In the interview with the mothers, the working mothers were asked, "How do you feel about working?" These answers were coded as to the predominance of positive or negative attitudes toward work. Responses to this question were on the positive side for most respondents. Sixty-five respondents were classified as liking work, and 23 as disliking it.

All hypotheses were tested separately for the two groups, i.e., by comparing the 65 mothers who liked work with their 65 matched nonworking counterparts and the 23 mothers who disliked work with their 23 matched nonworking counterparts. Significance tests for differences included the t test for correlated means where the data were quantitative, sign tests where the data were qualitative, and the t test for proportions where the data could not be analyzed by pairs.

Results

Table 1 summarizes the results dealing with the mother's affect and behavior toward the child. The first comparison deals with the child's perception of the mother as a source of positive affect. The child had been asked to indicate the persons who, from his own experience, best fitted each of a series of verbs, such as "praises" and "smiles," and to indicate the

TABLE 1

MOTHER'S AFFECT AND BEHAVIOR TOWARD CHILD: WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Child's report		
positive affect from mother	more than nonworking*	less than nonworking
coerciveness from mother	less than nonworking	less than nonworking*
Mother's report		
severity of discipline†	less than nonworking*	no difference
power assertion by the mother† ..	less than nonworking	less than nonworking*
mother's feelings of sympathy† ..	more than nonworking*	no difference
mother's feelings of hostility† ...	less than nonworking*	less than nonworking

NOTE.—Findings reported in all tables are based on the *t* test for significance of difference between correlated means except where indicated.

* Significant at the .05 level, one-tailed test.

† Comparisons not possible between matched pairs. Each working group therefore compared to pooled sample of nonworking mothers and the *t* test for significance of difference between proportions used.

degree of fitness of person to verb. The perception of positive affect in the mother-child relationship was measured by the extent to which the child associated "mother" with the following words: "praises," "smiles," "listens," "helps," "explains." As can be seen in Table 1, where the working mother indicates a positive attitude toward her work, the child associates more positive affect with the mother than do children in the matched group of nonworking women. This difference is significant. When we compare the children of working mothers who have negative attitudes toward their work to the nonworking group, we find the opposite relation although it is not a significant one.[2]

The second set of findings deals with the child's perception of coerciveness on the part of the mother. This variable was similarly measured using the words, "threatens," "punishes," and "hits." The table shows that the working mother is less likely than the nonworking mother to be associated with coerciveness in both groups and that the relation is significant for the working mothers who dislike work.

The remaining findings in this table are based on the mother's report. In the interview, mothers were asked to give detailed accounts of two recent occasions when they wanted the child to do something that the child did not want to do. These descriptive accounts were coded for degree of severity

[2] The direction of the relation is reported even where differences are small; so, except for the significant relations, differences can be due to chance variations.

of discipline, degree of power assertion over the child, sympathy, and hostility. Unfortunately, these descriptions were not available on all of the matched nonworking respondents. To have a large enough *N*, it was therefore necessary to pool the nonworking respondents and compare the two groups of working women separately with this pooled group. These are not, then, matched comparisons, although the working and nonworking groups are, on the whole, comparable with respect to the matching variables. The findings show that where the mother has positive attitudes toward employment she is less severe in her discipline, uses less power-assertive influence techniques, and feels more sympathy and less hostility during the interaction than is the case with the nonworking group. For the working mothers with negative attitudes toward employment there is less power assertion and somewhat less hostility.

Taking the findings in Table 1 as a whole, they tend to support the hypothesis that the working woman who gains pleasure from her work shows the child more affection and uses milder discipline. Although she is emotionally involved in the discipline situation, the involvement is to a great extent one of sympathy for the child. The working mother who does not enjoy her work, and whom we would expect to be relatively guilt-free thereby, also shows a tendency toward mild discipline. However, here it is not part of a larger pattern which includes affection and sympathy toward the child. Instead it seems to be part of a general pattern of less involvement with the child. This suggests that the dislike-work group tends to withdraw from the maternal role.

Table 2 deals with the child's participation in tasks. This was measured by the child's responses to a form of the "Doing Things" questionnaire

TABLE 2

CHILD'S PARTICIPATION IN HOUSEHOLD TASKS: WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Child's report		
regular participation in household tasks	less than nonworking	more than nonworking*

* Significant at the .05 level, one-tailed test.

developed by Herbst (6), in which the child is asked to report the extent of his participation in several household tasks (20 tasks in this case). As expected, the children of the working mothers who do not enjoy their work participate more in household tasks than do the children of the nonworking mothers; the children of the working mothers who enjoy work participate

TABLE 3

CHILD'S BEHAVIOR:
WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Mother's report		
child's assertiveness toward mother‡	no difference	more than nonworking*
child's playing with younger children rather than own age or older§	more than nonworking*	no difference
Teacher's report		
initiation of friendships	less than nonworking*	less than nonworking
influence attempts made to peers	less than nonworking	more than nonworking
influence success with peers	less than nonworking*	no difference
teacher dependency (girls)	no difference	no difference
teacher dependency (boys)	more than nonworking	more than nonworking†
aggressiveness	no difference	more than nonworking
use of physical force	less than nonworking	more than nonworking*
impulse control	no difference	less than nonworking*

* Significant at the .05 level, one-tailed test.

† Significant at the .05 level, two-tailed test, difference not predicted.

‡ Comparisons not possible between matched pairs. Each working group therefore compared to pooled sample of nonworking mothers and the t test for significance of difference between proportions used.

§ Sign test used for this variable.

less. This is in keeping with the notion that the mothers who enjoy work go out of their way to avoid inconveniencing their children because of their employment.

Table 3 deals with the child's behavior. The child's assertiveness toward the mother is based on the mother's account of the conflict situations in which she wanted the child to do something that the child did not want to do. The child's behavior was considered assertive when he directly indicated that he did not wish to comply and that he did not expect to do so. These behaviors included requests such as "Is it all right to stay up until the end of this program?"—as well as statements of intention such as "No. I am not going to bed until the end of this program."—where the context made it clear that the child was not "just talking." Pleading, crying, complying with complaints or silent anger, and devious attempts to obtain ends ("I have a stomach ache.") were all considered less assertive. The child's tendency to play with younger children is based on the mother's report that his playmates were "mostly older," "mostly younger," or "the same age."

The remaining variables reported in Table 3 are based on teacher ratings. These ratings are described more fully elsewhere (8).

It was hypothesized that the children of working mothers who have positive attitudes toward employment would be nonhostile, nonassertive, and possibly withdrawn and passive as compared with their nonworking counterparts, but that the children of working women who do not like work would be assertive and hostile. The data seem to bear this out. The children whose mothers like work differ in their general lack of initiative-taking and tend to play with children younger than themselves. The children whose mothers do not like work, on the other hand, show assertive behavior toward their mothers and toward their peers. This assertiveness toward peers includes aggressiveness, use of physical force, and less impulse control in general. Only teacher dependency does not operate as expected, the boys in both groups showing more dependency, significantly more in the dislike-work group.

Both the "guilt-overprotection" theory and the "neglect" theory lead to the prediction that the children of working mothers will be more disturbed in general than the children of nonworking mothers. This was tested by comparing the matched pairs on the use of nonadaptive responses to frustration, intellectual performance, and sociometric ratings by classmates. The results of these tests are reported in Table 4. The first measure was based on a teacher rating of how the child typically reacts to frustration. "Realistic

TABLE 4

CHILD'S ADJUSTMENT:
WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Teacher's report		
nonadaptive response to frustration†	more than nonworking	more than nonworking*
intellectual performance	less than nonworking*	less than nonworking*
Classroom sociometric		
liking by others	less than nonworking	less than nonworking

* Significant at the .05 level, one-tailed test.
† Sign test used for this variable.

acceptance of insolvable problem" and "attacks problem directly" were considered adaptive responses, while responses such as "cries," "goes to pieces," "blames others," and "becomes overcritical of self" were considered nonadaptive. Both groups of working-mother children were more likely to use nonadaptive responses to frustration. They also showed lower intel-

lectual performance and were somewhat less liked by the other children in the class.

The lower intellectual performance of children whose mothers work is a particularly interesting finding. Being low on performance may be a function of low ability or low motivation. If it is true that the working mothers who like work are guilty about their employment and consequently overprotect their children, then the intellectual *ability* of their children may be impaired. Mothers who solve their children's problems for them may hamper their intellectual development by depriving them of valuable problem-solving experience. Thus, these children would be expected to be low on ability as well as performance. The low performance of the dislike-work children, on the other hand, might be part of the rebellious pattern that seems to characterize this group; here we would not necessarily expect low ability. Scores on the Primary Learning Aptitude Test for third graders and the Detroit Alpha Intelligence Test for fourth, fifth, and sixth graders were available. The expectations were borne out. Only the children whose mothers were positive about their work obtained lower scores on these tests than their nonworking counterparts, and this difference was significant. The others showed no such difference.

Summary and Conclusions

The over-all pattern of findings suggests that the working mother who likes working is relatively high on positive affect toward the child, uses mild discipline, and tends to avoid inconveniencing the child with household tasks; the child is relative nonassertive and ineffective. The working mother who dislikes working, on the other hand, seems less involved with the child altogether and obtains the child's help with tasks; the child is assertive and hostile.

The data do not show conclusively that the reason for these different patterns is the presence or absence of guilt in the mother, but they are consistent with such an interpretation. If one were to argue that the causality is reversed, that the attitude toward work is a function of the effect of employment on the child or that happier women produce happier children, it would not be consistent with the fact that both groups of working-mother children show indications of maladjustment. It is only that they show different *syndromes* of maladjustment. Furthermore, in Table 5 it can be seen that the attitude toward employment is in large part a function of the particular job the woman holds.[3]

It is interesting to note in Table 5 that, while attitude toward employment is not purely a function of social class, there is somewhat of a tendency for higher status jobs to be well liked. Because middle-class women,

[3] The data presented in Table 5 are consistent with the data reported by Weiss and Samelson (12). Their data also show, for example, a positive attitude toward employment on the part of domestic workers.

TABLE 5

OCCUPATION AND ATTITUDE TOWARD WORK

Occupation	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Professional and semiprofessional, e.g., teacher, social worker, registered nurse	14	1
Skilled, e.g., secretary, bookkeeper	11	7
Sales	14	0
Waitress	1	4
Semiskilled or unskilled factory and clerical	20	10
Domestic	5	1
Total	65	23

compared to lower-class, can obtain higher status jobs, because their employment is likely to be choice rather than necessity, and because there is probably greater anxiey about childrearing in the middle class, it is likely that guilt and overprotection are more a middle-class response to maternal employment while neglect is more a lower-class response.

The real purpose of this paper, however, is not so much to elaborate on any particular theory as to suggest an approach to the study of the effects of maternal employment. The need to control on spurious variables has been established. The problem now is to differentiate the all-inclusive variable of maternal employment into more highly specified categories. Only in this way can we come to understand the processes by means of which employment affects the child. If one wants to know whether or not the sheer absence of the mother from the home is the important variable, then one should subdivide the working women according to the number of hours they are away from home while the child is there. If one wants to know the importance of the adequacy of supervision, then one should subdivide the women according to whether or not the substitute supervision provided is adequate. One might feel that the important variable is the father's reaction to the mother's employment. In this case, the working women can be subdivided according to some measure of the father's attitudinal or behavioral response to his wife's employment. In short, by selecting a particular aspect of mother's employment, differentiating the working mothers according to this variable, and examining the differences between working and nonworking groups separately, one can make a more formidable step toward understanding the crucial aspects of maternal employment in relation to the child.

There are still other advantages of this kind of approach. When the working-mother group is not differentiated, it includes a heterogeneous group and, as such, the working-nonworking comparisons will obscure

differences due to counteracting subgroup relations. For example, in the data reported here, since one group of working-mother children was less assertive and the other group was more assertive, the two tendencies canceled one another out. Thus, when the total group of working-mother children was compared with the matched nonworking group, there appeared to be no differences in child assertiveness. The same is true for the participation of children in household tasks.

Another kind of error that can result from failure to differentiate the working mothers is when what seems to be a single pattern of findings actually includes two or more distinct patterns. For example, when the total groups were compared, working mothers' children were found to play more with younger children and also to show less impulse control. Doing separate analyses for the two groups of working mothers, however, revealed that these results did not come from the same subjects. Although it was not true of the present data to any great extent, it would be possible for a pattern to emerge from the general working-nonworking comparisons which actually characterized neither the total group nor any single subgroup, but was entirely a coincidence of findings contributed by different clusters of subjects.

One final point should be made. In most working-mother studies, relations are examined between maternal employment and some aspect of the child's behavior *directly,* without considering intervening variables such as the effect of maternal employment on parent-child relations, family structure, or other aspects of family life. The jump between maternal employment and child behavior is too broad to be covered in one leap. One must first understand the various steps that intervene.

In summary, this paper suggests that maternal employment has a different effect on the mother-child relationship and on the child's behavior depending on whether or not the mother enjoys working. The purpose in doing this analysis was to show that greater understanding of the effects of maternal employment will come about when working mothers are differentiated along some theoretically relevant dimension and when the effects on family life, as well as on the child, are examined.

References

1. Bandura, A., & Walters, R. H. *Adolescent aggression.* Ronald Press, 1959.
2. Cyrus, D. Problems of the modern homemaker-mother. In J. T. Landis and M. G. Landis (Eds.) *Readings in marriage and the family.* Prentice-Hall, 1952. Pp. 392-402.
3. Glueck, S., & Glueck, E. Working mothers and delinquency. *Ment. Hyg.,* 1957, 327-352.
4. Hand, H. B. Working mothers and maladjusted children. *J. educ. Sociol.,* 1957, 30, 245-246.

5. Henry, A. F. Sibling structure and perception of the disciplinary roles of parents. *Sociometry*, 1957, 20, 67-74.

6. Herbst, P. G. Analysis and measurement of a situation: The child in the family. *Hum. Rel.*, 1953, 6, 113-140.

7. Hoffman, L. W. Effects of the employment of mothers on parental power relations and the division of household tasks. *Marriage Fam. Living*, 1960, 22, 27-35.

8. Hoffman, L. W., Rosen, S., & Lippitt, R. Parental coerciveness, child autonomy and child's role at school. *Sociometry*, 1960, 23, 15-22.

9. Nye, F. *Family relationships and delinquent behavior.* Wiley, 1958.

10. Rouman, J. School children's problems as related to parental factors. *J. educ. Res.*, 1956, 50, 105-112.

11. Siegel, A. E., Stolz, L. M., Hitchcock, E. A., & Adamson, J. Dependence and independence in the children of working mothers. *Child Develpm.*, 1959, 30, 533-547.

12. Weiss, R. S., & Samelson, N. M. Social roles of American women: their contribution to a sense of usefulness and importance. *Marriage Fam. Living*, 1958, 20, 358-366.

TEST ANXIETY AND CLASSROOM OBSERVATIONS[1]

KENNETH S. DAVIDSON* *and* SEYMOUR B. SARASON
Yale University

In a previous publication (4) we have described a variety of studies comparing high anxious (HA) and low anxious (LA) children. The measure primarily used for the assessment of anxiety level was the Test Anxiety Scale for Children (TASC). The pattern of these previous findings has been clearly and extensively less favorable for the high anxious children, particularly for the high anxious boys.

In one study (3) classroom observations were made of HA and LA children matched for grade, sex, and intelligence. Those observations described the HA children (particularly the boys) as less secure, task-oriented, and academically adequate than the LA subjects. Because those classroom observations were primarily exploratory and one of a series of procedures for studying 32 matched pairs of HA and LA children, each child was observed for only one hour. The data obtained from these observations were necessarily limited in scope and could not clarify the nature of dynamics and changes occurring in children that would aid in understanding the process by which anxiety interferes with scholastic effectiveness. In that study the process of subject selection did not permit assessment of the effects resulting from differences in teachers and classroom atmospheres.

As a result of these considerations the classroom observations in the present studies had two purposes: (a) to determine the relation of anxiety about school to a wide variety of personality and behavior variables measured in the classroom and (b) to explore the effects of differences in classroom atmosphere upon those relations.

PROCEDURE

The present study was carried out in the three second grade classrooms of a North Haven, Connecticut, elementary school. The choice of the second grade also permitted study of the important question of how early high

* Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education and a grant from the National Institute of Mental Health, Department of Health, Education, and Welfare.

Child Develpm., 1961, 32, 199-210.

anxious children can be detected in order to initiate early steps towards prevention of negative effects of anxiety on learning and development. The TASC can be administered in the usual group procedure to most children at the end of first grade. In the last month of the previous school year the TASC and a similarly constructed and administered questionnaire, the Defensiveness Scale for Children (DSC),[2] were given to the first grade children.

The DSC includes the 11 Lie items formerly found in the General Anxiety Scale for Children (GASC) (4) and 27 other items. Children who give "Yes" answers to questions of the DSC are saying in effect that they can admit and communicate to others about a variety of reactions such as anxiety, shame, aggression, insecurity, and concern about themselves and their relations with others. Children who answer "No" more frequently would generally be more "defensive." The interest in a "defensiveness" score was an exploratory one and stemmed from various considerations. For example HA children were able to admit to feelings of anxiety on the TASC and GASC, but parental interview findings (1, 4) indicated that they did not communicate freely about their feelings to their parents. Also, some LA children with high Lie scores seemed to communicate in a defensive manner that was more typical of HA children. Finally, the TASC discriminated very effectively between HA and LA boys but significant results were much less frequent for girls. The hope was that the DSC would differentiate among girls more effectively or aid in understanding the limited results obtained for girls with the TASC. For these reasons the defensiveness scores were obtained and correlated, in an exploratory manner, with the same measures of personality and classroom behaviors as were the TASC scores.

All of the children in the three second grade classrooms were observed by two trained observers who made notes spontaneously. They observed all morning or all afternoon daily for four months beginning the first day of school. In the original plan one observer sat in one room, the other in a second room. Both observed the children in the third room in order to provide a measure of the adequacy of inter-observer agreement in rating children on personality and classroom behavior variables. After a few observation sessions in this third room, both independently reported that the children expressed themselves much less freely in the third room than the children in the other rooms. Each felt they were not getting to understand the children in the third room as well, which indicated that, unless more freedom of expression occurred in those children, the inter-observer reliability in evaluations of those children could be inadequate for subsequent analyses.

As a result, beginning about the third week, both observers sat, at different times, in each of the three second grade rooms and took notes on all the

[2] The last item (4) in the references contains the TASC and the GASC as well as a chapter on adminstration of these scales. The DSC can be obtained from the authors for research purposes only.

children. The number of children in the study varied because families moved, but the maximum number of subjects in any aspect of the study was 96, including 53 boys and 43 girls, of whom 40 boys and 37 girls had received the TASC and the DSC in first grade.

In order to assess the subjects' personality and classroom behavior, a check list of 25 variables was formed with definitions provided for each of the items. After three months both observers rated all 96 children, and each teacher rated the children in her class on a 10-point scale of each of the 25 items. A number of the items were selected because they had yielded significant results when parents of HA and LA subjects had rated their own child on those qualities (2, 4). Other items dealing with daydreaming, level of achievement, speed of learning, memory, and attention to schoolwork were included because of their possible significance for both anxiety and classroom effectiveness. However, predictions were not made for these items because children's behavior in school may vary considerably from their behavior at home and because parents and teachers may differ in their values, expectations, and sanctions regarding children's behavior and personality.

Results

There were two phases in the analysis of the data obtained in this study. In the first phase the observers' and teachers' check-list ratings were analyzed to determine the degree of inter-observer and teacher-observer agreement. Also, the children's TASC and DSC scores were correlated with the check-list ratings given them by the observers and their teacher. In the second phase the observers' classroom notes were analyzed to compare the teachers' methods of handling children's needs and the teachers' expression of emotions and value judgments.

Personality Check List

Inter-observer reliability. The check-list study began with an assessment of the inter-observer reliability. The two observers' ratings of the children were correlated for each of the 25 items of the check list. These correlations were computed for the following six groups: all 96 subjects, the 53 boys, the 43 girls, and each of the three separate classrooms with boys and girls combined. The results of this analysis are presented in Table 1. All subsequent analyses involving check-list ratings utilized only those 18 items with inter-observer reliability levels of at least .40 for boys and girls both separately and combined. Two major findings emerge from this assessment of reliability: First, there is significantly greater agreement between observers in evaluating boys as compared with girls. On 21 of the 25 items, the reliability coefficient is higher for boys than for girls ($p < .01$). Second, the inter-observer agreement for the children in rooms 2 and 3 is similarly greater than for the children in room 1, which was the room originally chosen for the reliability check. It should also be noted that there are uni-

TABLE 1

INTER-OBSERVER RELIABILITIES*

Item	Children N=96	Boys N=53	Girls N=43	Room 1 N=32	Room 2 N=32	Room 3 N=32
1. Anxiety§	.27	.36	.23‡	.20‡	.28‡	.32†
2. Anxiety facilitates§	.48	.65	.08‡	.47	.68	.31†
3. Independent	.48	.51	.41	.29‡	.59	.54
4. Hides emotion	.61	.66	.56	.47	.67	.71
5. Difficulty in communicating	.49	.50	.53	.58	.60	.52
6. Submissive	.70	.79	.54	.60	.73	.80
7. Cautiousness	.69	.71	.58	.58	.66	.71
8. Not sensitive	.46	.50	.50	.13‡	.22‡	.78
9. Relaxed§	.23†	.33†	.12‡	.38†	.25‡	.09‡
10. Unambitious	.47	.44	.50	.30†	.53	.49
11. Set in ways§	.64	.79	.30†	.65	.59	.72
12. Not well liked	.56	.77	.42	.39†	.64	.67
13. Immature	.61	.63	.63	.57	.67	.58
14. Sociable§	.66	.75	.33†	.35†	.66	.85
15. Does not daydream	.47	.45	.50	.39†	.60	.42
16. Inactive	.62	.74	.46	.63	.73	.46
17. Underachievers	.49	.49	.50	.17‡	.76	.71
18. Learns quickly	.88	.74	.83	.74	.86	.82
19. Forgets	.78	.78	.75	.65	.83	.83
20. Does not fear failure§	.33	.31†	.10‡	—.09‡	.09‡	.52
21. Does not pay attention	.72	.87	.57	.65	.56	.81
22. Weak conscience	.73	.78	.55	.66	.70	.85
23. Masculine	.75	.66	.56	.51	.96	.78
24. Optimistic§	.46	.54	.39	.16‡	.60	.57
25. Not responsible	.78	.84	.62	.78	.74	.84
Mean (25 items)	.60	.65	.48	.44	.65	.67
Mean (18 items)	.65	.68	.57	.48	.70	.71

* All correlations are significant (one-tail test) beyond the .01 level except as noted.

† $p < .05$.

‡ Not significant.

§ These items were not used in later analyses or in computing means in the bottom row.

formly low levels of agreement between observers about children's anxiety, tension, and fear of failure in comparison to their very adequate agreement on variables such as emotional expression, cautiousness, maturity, speed of learning, retention, and responsibility.

Teacher-observer agreement. As part of this phase of the analysis, teacher-observer agreement was assessed by correlating the teachers' ratings of the children with those by each observer for each of the 18 selected items. These correlations were computed for the three separate rooms with the boys and girls combined in each room. The results are presented in Table 2.

TABLE 2

CORRELATIONS BETWEEN OBSERVERS' AND TEACHERS' CHECK-LIST RATINGS*

	Teacher 1 $N=32$		Teacher 2 $N=32$		Teacher 3 $N=32$	
	Rater A	*Rater B*	*Rater A*	*Rater B*	*Rater A*	*Rater B*
1. Independent	.13‡	—.06‡	.55	.80	.02‡	.12‡
2. Hides emotions	.16‡	.58	.30†	.47	.45	.45
3. Difficulty in communicating	.67	.60	.58	.73	.32	.60
4. Submissive	.67	.62	.74	.66	.70	.60
5. Cautious	.69	.37†	.59	.72	.76	.68
6. Not sensitive	.25‡	.36†	.50	.35†	.53	.59
7. Unambitious	.69	.52	.54	.32†	.60	.52
8. Not well liked	.52	.36†	.65	.54	.60	.22‡
9. Immature	.23‡	.34†	.61	.66	.78	.65
10. Does not daydream	.14‡	.50	.27‡	.49	—.04‡	.35†
11. Inactive	.53	.67	.78	.80	.71	.39†
12. Underachievers	.04‡	.45	.56	.64	.31†	.51
13. Learns quickly	.68	.65	.79	.84	.73	.78
14. Forgets	.44	.55	.86	.82	.67	.70
15. Does not pay attention	.69	.56	.74	.73	.80	.74
16. Weak conscience	.69	.61	.78	.81	.86	.79
17. Masculine	.83	.73	.91	.71	.68	.72
18. Not responsible	.67	.47	.79	.80	.81	.77
Mean	.52	.51	.68	.69	.62	.59

* All correlations are significant (one-tail test) beyond the .01 level except as noted.
† $p < .05$.
‡ Not significant.

The observer-teacher correlations are higher for room 2 than for room 1 for 15 of the 18 items in the case of each observer. Comparing rooms 3 and 1 the correlations are larger in room 3 for 13 of the 18 items in the case of observer A and larger for 10 of the 18 items for observer B. The means of the 18 correlations for each teacher-observer combination show that, for each observer, agreement with the teachers' ratings was greatest for room 2, smaller for room 3, and least for room 1. The difference in mean teacher-observer agreement between room 2 and room 1 was significant for both observers ($p < .01$). Thus, not only the inter-observer reliability but also the teacher-observer agreement is significantly less for the check-list ratings of the children in room 1.

It will be recalled that both observers remarked very early in their observations that the children in room 1 seemed to express themselves less freely than the other boys and girls. This suggested that the observers' and teachers' ratings might be more constricted for the children in room 1 as compared with rooms 2 and 3. Thus, for each item, the standard deviations were com-

puted for each observer's and the teacher's ratings of the children in the separate rooms. Fourteen of the standard deviations for observer A are larger than the corresponding sigmas obtained from the teacher's ratings in room 1. For observer B, 13 of the 18 standard deviations are larger than those obtained for this teacher. However, the outstanding finding is obtained by comparing the standard deviations of the ratings by the three teachers. For all 18 items, the teacher in room 1 constricted her ratings more than the other two teachers and often did so to a marked degree. While the observers tended—observer A only slightly—to constrict their ratings of the children in room 1 more than the other children in the other rooms, the teacher in room 1 constricted her ratings of her pupils more than either observer and much more than the other two teachers.

Relationship of ratings to test anxiety and defensiveness. The last phase of the check-list analysis concerned the relation between the observers' and

TABLE 3

SIGNIFICANT MEAN CORRELATIONS BETWEEN SUBJECTS' TASC SCORES AND RATINGS BY OBSERVERS AND TEACHERS

*Variable**	*Room 1* $N=27$	*Room 2* $N=24$	*Room 3* $N=26$	*Boys* $N=40$	*Girls* $N=37$	*Children* $N=77$
1. Independent	..	..	..	+.24‡	−.26‡	..
2. Hides emotions	..	+.25†	..	+.27‡	−.19†	..
3. Difficulty in communicating	..	..	..	+.20†	..	..
4. Submissive	..	+.54‡	..	+.23†	..	+.15†
5. Cautious	..	+.47‡	..	+.30‡	..	+.14†
6. Not sensitive	..	−.30†	..	..	−.28‡	−.16†
7. Unambitious	..	..	..	−.20†	..	..
8. Not well liked	..	..	..	..	..	..
9. Immature	..	..	..	..	+.22†	..
10. Does not daydream	..	..	..	..	..	+.14†
11. Inactive	..	+.34‡	..	+.25‡	..	..
12. Underachievers	..	−.26†	..	−.25‡	..	−.17†
13. Learns quickly	..	..	+.31‡	..	..	..
14. Forgets	..	..	+.30†	..	..	..
15. Does not pay attention	..	−.47‡	..	−.26‡	..	−.18‡
16. Weak conscience	..	−.58‡	..	−.26‡	..	−.21‡
17. Masculine	..	−.45‡	+.24†	..	..	−.16†
18. Not responsible	..	−.39‡	..	−.29‡	..	−.14†

NOTE.—Mean correlations were computed through z transformations and tested for significance using a two-tail test and the formula, $t_z = M_z\sqrt{3(N-3)}$, with $df = 3(N-3)$.

* Positive correlations mean that anxiety is directly related to the variables as listed, while for negative correlations anxiety is inversely related to the listed variables.

† $p < .05$.

‡ $p < .01$.

teachers' check-list ratings of the children and the subjects' TASC and DSC scores. For the children in each room and for the boys and girls separately and combined in all rooms, TASC and DSC scores were correlated with each observer's and the teacher's ratings. For each grouping of children, three correlations for each item were obtained between ratings—the two observers' and the teachers'—and the TASC scores. The means of these three correlations which were significant are presented in Table 3. The same procedure was followed for the DSC, and these results are presented in Table 4.

TABLE 4

SIGNIFICANT MEAN CORRELATIONS BETWEEN SUBJECTS' DSC SCORES AND RATINGS BY OBSERVERS AND TEACHERS

*Variable**	*Room 1* $N=27$	*Room 2* $N=24$	*Room 3* $N=26$	*Boys* $N=40$	*Girls* $N=37$	*Children* $N=77$
1. Independent	+.37‡	..	−.26†	..	..	..
2. Hides emotions	+.34‡	−.32‡	..	..	..	..
3. Difficulty in communicating	+.36‡	..	..	..	+.28‡	+.16†
4. Submissive	+.40‡	−.31†	..	..	..	..
5. Cautious	+.54‡	−.28†	−.33‡	..	..	..
6. Not sensitive	+.35‡	..	..	+.22†	..	..
7. Unambitious	..	..	+.36‡	..	..	+.17†
8. Not well liked	..	..	..	..	+.27‡	+.14†
9. Immature	−.40‡	..	+.50‡	..	..	..
10. Does not daydream	..	..	..	..	−.23†	..
11. Inactive	+.26†	..	..	..	..	..
12. Underachievers	..	..	+.46‡	..	+.20†	..
13. Learns quickly	..	..	−.27†	..	−.26‡	−.22‡
14. Forgets	..	..	+.43‡	+.19†	+.32‡	+.25‡
15. Does not pay attention	..	..	+.58‡	..	+.20†	+.18‡
16. Weak conscience	−.42‡	..	+.54‡	..	..	+.13†
17. Masculine	..	..	..	..	+.23†	+.22‡
18. Not responsible	−.31‡	..	+.54‡	..	+.23†	+.17†

NOTE.—Mean correlations were computed through z transformations and tested for significance using a two-tail test and the formula, $t_z = M_z\sqrt{3(N-3)}$, with $df = 3(N-3)$.

* Positive correlations mean that defensiveness is directly related to the variables as listed, while for negative correlations defensiveness is inversely related to the listed variables.

† $p < .05$.

‡ $p < .01$.

Aside from the specific relations between anxiety or defensiveness and the characteristics and behaviors measured by the ratings, two findings are clearly shown. First, TASC scores relate significantly to a whole series of personality characteristics for boys, but not for the girls, while the reverse is true for the DSC. Second, in room 2 anxiety correlates significantly with a

number of variables but not defensiveness. The reverse is true in both rooms 1 and 3. The first finding indicates that anxiety about school is related to a variety of personality traits and behavior patterns among boys but only to a few such qualities among girls. On the other hand, high defensive (HD) and low defensive (LD) girls differ in many respects while HD and LD boys tend not to differ much. In other words, anxiety level, as measured by the TASC, is an effective predictor of boys' personality and behavior patterns in school, whereas defensiveness, as measured by the DSC, performs in the same manner for girls.

In view of the variation in both inter-observer and teacher-observer agreement on the check-list ratings, one would expect corresponding results in correlating those ratings with both the TASC and DSC scores. However, in room 1, where there was the least agreement about ratings, the DSC yielded significant results for 10 of the 18 items and no significant results for any of the TASC items. In room 2, where there was greatest agreement about ratings, the TASC yielded significant results for 10 of the 18 items, but the DSC yielded significant results for only three items. In other words, in spite of the lower reliability for ratings of children in room 1, those ratings are significantly correlated with defensiveness. These findings suggest that either anxiety or defensiveness may be the meaningful variable for different classrooms.

Classroom Observation Notes

This phase of the study was concerned with the teachers' methods of coping with the children's needs and personality characteristics. The analysis focused on teachers' handling of children's dependence, cautiousness, fear of failure, motivation, failure and success, and behavior problems (shyness, anxiety, aggression). Instances in which the teacher expressed strong emotions or value judgments were also noted.

The observers independently culled examples from their notes for each teacher for each of these categories. For each incident or note, two clinicians independently indicated by favorable (+) or unfavorable (−) ratings whether the teachers' efforts had helped the children by meeting their needs or had interfered with or reduced the children's effectiveness. Some notes were rated as favorable with unfavorable tones (+ −) or as more unfavorable (− +) while others were given ?'s when both qualities were equal. The clinicians agreed completely on 628 or 70 per cent of their 892 ratings, while disagreeing completely on less than 4 per cent of the ratings. On 745 or 84 per cent of their ratings they agreed about the favorable or unfavorable direction of the rating.[3]

With this extensive agreement the two clinicians held conferences to arrive at a single rating for those notes on which they disagreed. Thus, for

[3] The authors wish to express their appreciation to Dr. Wayne Owen for his assistance in these detailed evaluations. In addition the reader may wish to refer to the parental interview study (1; 4, pp. 222-225) in which the same general procedure of "favorable-unfavorable" ratings was used.

TABLE 5

FREQUENCIES OF POSITIVE AND NEGATIVE RATINGS OF OBSERVERS' NOTES

	Teacher 1		*Teacher 2*		*Teacher 3*	
	+	−	+	−	+	−
Dependency	18	72	98	16	37	31
Caution	5	29	40	9	17	16
Fear of failure	5	11	24	4	13	7
Motivation	11	31	24	5	16	19
Failure and success	16	37	40	1	29	13
Shyness, anxiety, aggression	7	10	29	2	12	26
Emotion and value judgment	2	33	28	7	2	14
Total	64	223	283	44	126	126

each note there was one rating on which the judges were in agreement. The frequencies of favorable (+ and +−) and unfavorable (− and −+) ratings for each variable are given for the three teachers in Table 5.

For all seven categories, the teacher in room 2 received more favorable evaluations by far than either of the other two teachers. For the teacher in room 3, one half of the incidents were rated favorably and one half unfavorably, while only one quarter of the ratings for the teacher in room 1 were favorable. This pattern of results is especially clear with respect to the three teachers' handling of the children's dependence, cautiousness or impulsiveness, and failures or successes, as well as the teachers' expression of emotions and value judgments.

DISCUSSION

While these analyses yielded a series of significant results, there are really two general findings of importance for the study of child behavior and personality in the classroom. First, the evidence indicates that the teacher is a prime variable in such studies, and, second, different patterns of results can be expected for boys as compared with girls.

With respect to the first point, differences were found between teachers with respect to levels of reliability and constriction of check-list ratings, correlation of those ratings with TASC and DSC scores, and the evaluations of observers' notes on the teachers' methods and behavior. In the case of room 1 there was low inter-observer and teacher-observer reliability, constriction, especially by the teacher in rating the children, and unfavorable evaluations of the teacher's behavior. In comparison, for rooms 2 and 3, there was greater inter-observer and teacher-observer agreement and less constric-

tion in the rating of those children, as well as more favorable evaluations of the teachers.

It is distinctly possible that in room 1 the teacher's ineffective handling of the dependence, cautiousness or impulsiveness, failures and successes of her pupils, as well as her negatively evaluated expressions of affect and value judgments, were important determinants of the lack of spontaneous expression and behavior of those children. The notes of each observer and even casual or untrained observations describe this teacher as authoritarian, punitive, and rejecting of needs to be dependent or cautious. As a result, these children did conform to their teacher's rules and suppressed expression of their needs and feelings in a variety of ways.

It should be pointed out that the mean TASC scores were nearly identical for the three rooms. However, the mean DSC score for room 1 was higher, though not significantly, than the means for the other two rooms. Also, the mean Lorge-Thorndike IQ was lowest for room 1, though the difference was not significant. There is the possibility that the greater defensiveness and lower intelligence, in combination, accounts for their less spontaneous expression. It is our own opinion, however, that the greater defensiveness in room 1 may be an effect of the tendencies of the teacher to de-emphasize differences among her children (constriction of ratings) and to misperceive or misinterpret their behavior. That is to say, a teacher with such tendencies may not only be responding ineffectively and inappropriately to the needs and behavior of the children but also may be engendering in them greater defensiveness, i.e., the tendency to suppress the expression of personal feeling. We are not dealing here with a type of teacher-child interaction which occurs infrequently in this particular classroom, but one which goes on over long periods of time. It is also likely that, as this pattern of interaction goes on over time, the defensive behavior of the children affects the teacher in a way so as to reinforce her tendency to misperceive and misinterpret the significance of the behavior of her children.

The final difference between teachers showed defensiveness but not test anxiety related to a whole series of check-list items for the children in rooms 1 and 3, while the reverse was true for room 2. It is possible to speculate on the reasons for this pattern, but, in any event, it suggests that teachers who vary in their methods and behavior may stimulate different feelings and needs in their pupils which then determine their classroom behavior and characteristics. In rooms 1 and 3, receiving expressions of rejection or acceptance, respectively, seemed a central issue for the children, while in room 2 the teacher showed acceptance of all children, though not all behavior, and the central issue seemed to be academic achievement and improvement.

In view of the fact that anxiety was the meaningful variable for boys and defensiveness for the girls, it is possible that an important determinant of the similar result found among the teachers is differential reactions and relations to boys and girls by the teachers. Observers frequently commented that the teacher in room 1 tended to favor girls and the room 3 teacher

tended to favor boys, while there was no apparent difference for boys and girls in room 2. In any event, when teachers differ markedly, it seems highly likely that measures of different variables will yield different results for the children in their rooms.

The content of the above paragraph is but an instance of the more general finding of differential results for boys and girls. There was greater inter-observer and teacher-observer agreement in rating boys on the check-list items, and the observers and teachers did not constrict their ratings of the boys as much as they did those of the girls. Since both observers and all three teachers were women, it might be expected that, if anything, they would show more agreement and less constriction of their ratings for girls. That the reverse occurred is probably related to the fact that the observers and teachers agree that, in school, boys give more overt and spontaneous expression to a wider range of feelings and impulses than do girls. This finding does suggest that observation of children as a basis for measurement is likely to be more effective for a wider range of variables for boys than for girls.

The main difference between boys and girls, however, was the finding that test anxiety scores serve as predictors of boys' but not girls' school behavior and characteristics, while the pattern was reversed for the defensiveness scores. The fact is that differences between HA and LA have been found much more frequently for boys than for girls, and these results have generally showed the LA boys in a more favorable light. This led to speculation that feeling and admitting to anxiety is an ego-alien and disturbing state of affairs for boys, but may be ego-syntonic, not disturbing, and perhaps helpful in a motivational sense for girls (3). There are some other observations which bear upon this notion. Frequently, in giving the questionnaires, one sees girls nodding delightedly as they answer "Yes" to TASC items. Boys express pleasure in this situation only if they are shaking their heads sideways as they answer "No." Boys who answer these items "Yes" seem quiet and tense and often smile and express relief when they can give a "No" answer. These behaviors clearly fit in with the notion that being test anxious can be ego-alien for boys and ego-syntonic for girls.

There is no clear-cut behavior difference between boys and girls when they are responding to the DSC. However, this scale generally measures freedom to express feelings of anxiety, shame, aggression, and concern about the feelings of others when such expression is appropriate. Responding "Yes" means to the LD—in one sense—that their behavior is appropriate or "good" (i.e., concerned about others, accepting responsibility for oneself, etc.). "No" answers describe one as being able to cover up feelings or unconcerned about being aggressive or others' reactions—in a sense—a picture of a strong, silent, masculine person. Clearly, behaving in a "good" or appropriate manner would be ego-syntonic for both boys and girls. However, being unconcerned about others' feelings, masculine, or aggressive would be ego-syntonic for boys and ego-alien for girls.

While these points remain in the realm of speculation, nevertheless, they do indicate that the TASC may well measure ego-alien feelings for boys and ego-syntonic feelings for girls while the DSC measures ego-syntonic qualities of boys and ego-alien qualities of girls. If boys or girls can accept equally being high or low with respect to a given variable, that characteristic may be relatively meaningless as a determinant of their behavior and personalities. However, being high or low on a trait that is unacceptable is the difference between having a lot or a little of a disturbing quality or feeling. Anxiety, apparently ego-alien for boys and ego-syntonic for girls, is an effective predictor for the former but not the latter. Defensiveness, which may well measure ego-syntonic qualities for boys and ego-alien traits for girls, is an effective predictor of the behavior and personality characteristics of girls.

References

1. Davidson, K. S. Interviews of parents of high anxious and low anxious children. *Child Develpm.*, 1959, 30, 341-351.
2. Davidson, K. S., Sarason, S. B., Lighthall, F. F., Waite, R. R., & Sarnoff, I. Differences between mothers' and fathers' ratings of low and high anxious children. *Child Develpm.*, 1958, 29, 155-160.
3. Sarason, S. B., Davidson, K. S., Lighthall, F. F., & Waite, R. R. Classroom observations of high and low anxious children. *Child Develpm.*, 1958, 29, 287-295.
4. Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. *Test anxiety in elementary school children: a report of research.* Wiley, 1960.

IMITATIVE COLOR CHOICES IN KINDERGARTEN CHILDREN

JUDY F. ROSENBLITH*[1]
Brown University

Miller and Dollard (2) distinguished several types of imitative behavior among which is one they call *copying* and one they call *matched-dependent behavior*. Copying refers to instances in which one person learns to model his behavior on that of another, while matched-dependent imitation refers to instances in which (a) the response of the imitator matches that of the leader and (b) the imitator is dependent on the leader for cues to act or to match behavior. Miller and Dollard assert that both copying and matched-dependent imitation are learned in the same way. Accordingly, one might expect both types of imitative behavior to be related to experimental manipulations in the same way.

The present study deals with matched-dependent imitation as a function of the same set of variables (or experimental conditions) that were related to copying behavior in an earlier report (5). The variables are: (a) the sex of the leader, (b) the sex of the leader in relation to that of the child, and (c) the way in which the leader interacts with the child. These variables were chosen for their relevance to both identification and learning theories.[2]

---

* Psychology Department, Brown University, Providence 12, Rhode Island.

[1] This paper is based on a dissertation submitted to Harvard University in partial fulfillment of the requirements for the Ph.D. The author would like to express gratitude to Dr. Wesley Allinsmith for his encouragement and counsel. This research was aided by a grant to the Laboratory of Human Development from the U. S. Public Health Service.

[2] A fuller discussion of historical and theoretical issues which led to this choice of variables is to be found in earlier reports (4, 5).

Child Develpm., 1961, 32, 211-223.

The design employed in the over-all study permits us to examine both copying, discussed previously (5), and matched-dependent imitative behaviors, not only in relation to the same experimental manipulations, but within the same subjects. Copying was assessed by the amount of improvement in performance on the Porteus Maze Test of Intelligence. This task allowed us to categorize the children on the basis of their level of pretest maze performance. Thus, we had a rough way of controlling for intelligence or mental age, which may be expected to affect a child's ability to benefit from a leader (or model) or may be related to differential altertness to cues provided by a model.

Since we assess the occurrence of matched-dependent imitative behavior (specifically, use by the child of the same *color* of pencil as that used by the model) in a situation focused on other behaviors (namely, solving the maze), we maximize the probability that the child will not perceive color matching as an instrumental response. The design furthermore enabled us to avoid the chance that the model would use the child's preferred color of pencil, since this preference had already been determined at the same time as the level of maze performance had been measured. As a matter of fact, the model's color choices were the exact reverse of the preference pattern shown by the child.

In this experiment, unlike those reported by Miller and Dollard (2), Schein (6), and Schwartz (7), we are not concerned with teaching the child to imitate. Here, even more completely than in the copying part of the experiment (5), we are relying on the fact that, as a result of their past experiences in the processes of socialization, kindergarten children will have learned to make imitative responses. Miller and Dollard pointed out that the tendency to match behavior becomes a secondary drive, a drive to imitate. We are counting on the presence of such a drive. The child is not rewarded for imitating the color choice of the model. The experimenter makes no comment on the color choice of either model or child. Thus, the question of color choice is not brought to the child's attention in any way. As a consequence of all these factors, we may be quite confident that, if imitative color matching occurs, it does not stem from rewards operating in this experimental situation, but from (a) a learned drive to imitate or (b) from generalization from previous learnings or (c) from an established identification with adults (the models used in this experiment).

Questions concerning identification theory that were raised previously (5) are particularly relevant here. In general, one would expect that, the more a child is identified with an adult, the more tendency the child would have to imitate responses of the adult, such as the adult's choice of colored pencils. Therefore, we might expect to find differences in the amount of imitation exhibited as a function of the sex of the leader and/or as a function of the leader's sex in relation to the child's sex. The role of "threat" is relevant in this experiment, if it is relevant to strength of identification. Therefore, in half the cases in the interaction between adult and child, the

adult withdrew attention from the child before the imitative behaviors were measured.

This paper will examine not only the relation of matched-dependent imitation to the above mentioned variables, but will also address itself to the question of whether both copying (learning by imitation) and matched-dependent (color matching) imitative behaviors follow the same pattern in response to experimental manipulations. In addition, we shall examine the possibility that learning by imitation of the kind that we might designate as task-oriented (or "instrumental" imitation) may be relatively more sensitive to "threat" than matched-dependent imitation (or "role" imitation) which may prove more sensitive to the reward of attention.

Method

The *Ss* were 80 kindergarten children from two public schools in a suburb of Boston composed of middle and upper-middle class residents.[3] Each child was led individually from the classroom to the experimental room by the experimenter in order to "play a game." When they arrived, *E* said to the child: "Here is a chair for you, and here is one for me (at the opposite sides of the table). Here are *some pencils* for you. This is a paper and pencil game. The first thing I want you to do is take *one of your pencils* and draw. . . ." There were four pencils in front of the child: one red, one blue, and one green, each producing its color, and a yellow one which made ordinary lead marks. Nothing further was said about the pencils unless the child asked a question. If he did, he was again told that the pencils were all there for him and he could use whichever ones he wanted.

The child was then administered the Porteus Maze Test (3); slight modifications in instructions and procedure were made to render them more suitable to the age of the children.[4] On the basis of their performance in this initial session, children were assigned to one of four groups. Group 1 consisted of those who passed only the first two mazes (i.e., the 3- and 4-year mazes). Group 2 consisted of those who passed a total of three or four of the mazes. Group 3 passed five mazes, and group 4 passed six or more of the mazes. We might note the fact that level of initial performance on the Porteus Maze Test of Intelligence should reflect levels of ability or maturity. Children from each group were then assigned to each of the treatments. In order to have larger *Ns* for comparisons, groups 1 and 2 will often be considered together in relation to groups 3 and 4.

One to three weeks after this initial session the child was brought from his room by the *E* to have "a second turn at the game." Just before arriving at the experimental room, the *E* said to each child: "You know what? I have

[3] These are the *Ss* who were in one of the treatment groups that had a leader in the previous study (5).

[4] Description of the exact procedure can be found in (4).

someone else here to play the game with us." *E* and child then entered the room, and *E* introduced the child to the other adult by first names. *E* then explained to the child that she (*E*) had something to do for a while and invited the child and the other adult to go to another table in the room and play with the toys on it until *E* was ready to play the game.[5] The child was seated facing the table of toys and the adult at the side of the table next to the child in such a way as to face both the table and the child.

It should be noted that *E* essentially was treating the other adult in the same manner as the child. *E* did nothing to enhance the adultness (or maleness or femaleness) of the leaders except to reward their proficiency in performance. They were not called Mr. or Mrs. and not described as fathers or mothers. No attention was called to their choice of pencil color. Each leader interacted with the child, however, as an adult would and not as another child would.

The different treatment conditions were: I. Male leader, child of the same sex, leader attentive to child throughout the 10-minute period that preceded the copying session proper. II. Male leader, child of the same sex, leader attentive to the child for five minutes, then leader said: "I'm sorry, I can't play with you any more, I have to read a book." Leader then turned his chair so that his back was toward the child and read a book for five minutes. III. Male leader, child of the opposite sex, attentive throughout (as in I). IV. Male leader, child of the opposite sex, withdrew attention after five minutes (as in II). This entire design was then repeated using a female leader.

At the end of the 10-minute period *E* said: "I'm ready to play the game now." Adult leader and child then took adjacent seats at the table on the side opposite *E* and facing her. Both the adult and the child had four colored pencils at their places as the child had had in the previous session. *E* said to the child: "You've played this game with me before. ——(adult) wasn't here when we played then, so how would it be if you take turns and give ——(adult) the first turn." The adult was then given the first maze which the child failed previously and the same instructions as were given the child. The leader picked up a pencil (of the color least used by the child in the initial session) and did the maze correctly at a rather slow pace. The child watched during this period and then was given a turn on the same maze. There were two such trials on each maze which the child had initially failed. After the third trial the adult shifted to the second least used color and continued to shift after every three trials. This was continued until the criterion for discontinuing testing was reached. The proportion of all trials on which the child used the same color of pencil as the adult provides the basic data. These data are shown in Table 1 which also depicts the design of the experiment.

[5] The toys were selected to have relevance to the questions of identification. They included more or less matched "male" and "female" items: make-up and shaving sets, carpenter and kitchen tools, doll family, and others.

TABLE 1

PERCENTAGE OF SAME COLOR CHOICES UNDER THE VARIOUS CONDITIONS

Groups	Boys		Girls	
	Attention	*Withdrawal*	*Attention*	*Withdrawal*
Male Leader				
1	10	11	50	28
2	35	26	54	20
3	33	60	79	13
4	40	66	14	39
All groups	29	37	50	24
Female Leader				
1	20	7	52	54
2	29	14	50	13
3	37	66	31	31
4	19	29	25	33
All groups	27	26	41	28

RESULTS

Sex of Subjects

We shall report on sex differences according to both sex of subject and sex of leader. For girls, 36 per cent of all color choices were imitative; for boys, 30 per cent ($p < .10$).[6] The differences in amount of imitation shown by boys and girls depended on the initial level of performance in part. Girls showed more color matched choices than boys ($p < .001$) in groups 1 and 2 (the poorer, less bright or able groups on initial level of performance). Boys showed more color matched choices than girls ($p < .07$) when initial level of maze performance placed them in the two top groups (3 and 4). The over-all tendency of the girls to show more color matching than boys occurred both with the male leader and with the female leader, but was not significant with either. This nonsignificant tendency for girls to exhibit more color matching than boys is in contrast to the greater learning-by-imitation of the boys who achieved 101 new passes on the mazes to the girls' 69 (*see* Table 2).

Sex of Leader

Thirty-five per cent of all color choices were the same as those of the male leader; 30 per cent were the same as those of the female leader. This difference is significant at the .20 level only. Here again, initial level of maze

[6] Probability levels are based on the critical ratio or z test.

TABLE 2

SEX DIFFERENCES IN PERCENTAGE OF IMITATIVE COLOR CHOICES

Groups	*Boys*	*Girls*	*p**
1 & 2	21	40	.001
3 & 4	43	31	.07
All	30	36	.10

* In this and subsequent tables the probability level is determined using the critical ratio or *z* test.

performance made a difference. The male leader elicited more color matching than the female leader from the brighter *Ss* (p is .07 for *Ss* whose initial level of performance put them in groups 3 or 4 and .04 for those in group 4); the brighter the *Ss*, the smaller the p value (*see* Table 3).

Level of Initial Performance on the Mazes

The interaction of this variable with the other variables was examined. An analysis of variance by ranks, the expansion of the Kruskal-Wallis test used in (5), showed that only the interaction between levels and sex of subject was significant ($p < .01$).

Attention and Withdrawal of Attention

For girls (with both male and female leaders) attention from the leader resulted in more color-matching behavior than did withdrawal of attention. This was true across all levels of performance ($p < .002$) (*see* Table 4). (*See also* Table 5). For *all* boys (with both male and female leaders) there was no significant tendency for attention or withdrawal to be more effective. This result reflects two opposing tendencies. Boys whose initial level of performance was highest (groups 3 and 4) showed a significant ($p < .02$) tendency for withdrawal of attention to be more effective than attention throughout. On the other hand, boys whose initial level of performance

TABLE 3

PERCENTAGE OF SAME COLOR CHOICES WITH MALE AND FEMALE LEADERS

Groups	*Male Leader*	*Female Leader*	*p*
1 & 2	30	29	*ns*
3 & 4	43	32	.07
4	40	23	.04
All	35	30	.20

TABLE 4

PERCENTAGE OF IMITATIVE RESPONSES IN RELATION TO ATTENTION AND WITHDRAWAL OF ATTENTION

Groups	A*	WA*	*p*	A*	WA*	*p*
1 & 2	24	16	.17	52	27	.003
3 & 4	32	54	.02	38	25	.16
All	28	32	*ns*	46	26	.002

* In this and subsequent tables, A stands for attentive throughout, and WA for withdraws attention half the time.

placed them in groups 1 and 2 showed a nonsignificant tendency for attention to be more effective.

For all boys with the female leader, there was no effect of attention versus withdrawal of attention. There was a slight tendency ($p = .18$) for boys in groups 3 and 4 combined to do more color matching when the female leader had withdrawn attention. For all boys with the male leader there was a nonsignificant tendency for withdrawal of attention to be more effective. The superiority of withdrawal of attention has, however, a p value of less than .05 for boys in groups 3 and 4 (*see* Table 5).

Sex of Leader Relative to Sex of Child

Boys showed more imitation of the leader of the same sex while girls showed more imitation for an attentive leader regardless of sex. Taking boys and girls together, we find that the leader of the opposite sex who was attentive elicited more color matching ($p < .03$) than the leader of the opposite sex who withdrew attention. The leader of the same sex who was attentive did not differ from the leader of the same sex who withdrew atten-

TABLE 5

PROBABILITIES* THAT DIFFERENCES BETWEEN TREATMENTS COULD BE OBTAINED BY CHANCE

	Male Leader				Female Leader			
	Boys		*Girls*		*Boys*		*Girls*	
Groups	A	WA	A	WA	A	WA	A	WA
1 & 2	*ns*		.01		.11		.10	
3 & 4		.05	.10			.18	*ns*	
All		*ns*	.002		*ns*		.14	

* Values are placed under the treatment that was more effective.

tion, but each of them tended to elicit more color matching ($p = .12$ and .15) than did the opposite-sexed leader who withdrew attention (*see* Table 6).

TABLE 6

EFFECTIVENESS OF TREATMENTS WITH SEX OF LEADER CONSIDERED RELATIVE TO SEX OF CHILD

Conditions Compared		*p*
OS* - A	with OS - WA	.03
SS* - WA	with OS - WA	.15
SS - A	with OS - WA	.12

* In this and subsequent tables, OS stands for leader of opposite sex to child and SS for leader of same sex.

The rank order of effectiveness of treatments for enhancing this color-matching form of imitation is shown in Table 7. Figures permitting comparison with the rank order of effectiveness of treatments for learning by imitation are also given.

TABLE 7

RANK ORDERS OF EFFECTIVENESS OF TREATMENTS FOR COLOR MATCHING AND FOR LEARNING-BY-IMITATION*

	SS - A		SS - WA		OS - A		OS - WA	
Boys	2	(3)	1	(1.5)	3.5	(1.5)	3.5	(4)
Girls	2	(3)	3	(4)	1	(1.5)	4	(1.5)

* The rank order for learning by imitation is given by the numbers in parentheses.

Number of Pencils Used

A form of behavior other than color matching can be looked at in connection with the influence of the adult model, namely, how many different pencils did the child use. Of the 80 children, 41 used the same color on all or all but one of the pretest trials. Only 15 children did so in the copying session. Every one of these 15 was a child who had used only one color in the pretest session, but only four of them used the same color they had used in the pretest.[7] Of the 15 who still used only one color, 11 had a leader who withdrew attention. Of the 26 who gave up using only one color, 17 had a leader who had been attentive throughout. Of the nine who gave up

[7] The number of children who used each of the colors exclusively in the pretest are: red 16, regular 14, blue 10, and green 1.

using only one color *and* who had a leader who withdrew attention, seven had a male leader. Attention leads to giving up the use of one color only; withdrawal of attention by a male leader may lead to the use of several colors of pencils.

A χ^2 test can be used to examine the relation between giving up one color and attentiveness of the leader; χ^2 (corrected for continuity) for the 2 by 2 table is 4.26, which for 1 degree of freedom has a p value of .04. Whether the leader was male or female and whether the leader was of the same or opposite sex were, however, not significant factors.

Discussion

It is interesting to compare these data on color matching with those on learning by imitation (4, 5). Hence, the main task of our discussion will be to place the two sets of results in apposition.

1. Sex of subject: Boys tended to show more learning by imitation and girls tended to show more color matching (matched-dependent or "role" oriented imitation).

2. Sex of leader: Whereas the male leader was more effective in producing learning by imitation, there was only a tendency for the male leader to evoke more same-color pencil choices.

3. Attention vs. withdrawal of attention: For girls there was a greater tendency for both kinds of imitation to occur when *Ss* had attention than when the leader withdrew attention. In the case of girls with the male leader, the attentive treatment produced more same color pencil choices than did withdrawal of attention. Attention also produced more learning by imitation. However, in the case of learning by imitation the male leader who withdrew attention from the girl was the second most effective condition (essentially tied with the most effective), whereas for color choices this treatment yielded the smallest percentage of same color choices.

For boys the male leader's withdrawal of attention produced a tendency toward more learning by imitation and more same color choices than did attention throughout. However, the female leader who was attentive was more effective than the female leader who withdrew attention, both for learning by imitation and color matching. The female leader who withdrew attention was the least effective of all treatment conditions for both learning by imitation and color matching.

For both sexes of *Ss* attention was more effective than withdrawal of attention in leading the child to give up the use of only one color of pencil.

4. Leader of the same or opposite sex to the child: Boys showed more color matching for a leader of the same sex regardless of treatment, but best performance in learning by imitation was for the same sexed leader who withdrew attention and the opposite sexed leader who was attentive. Girls showed more color matching (as indicated in 3 above) for an attentive leader regardless of sex. On the other hand, they had shown more learning

by imitation for the opposite sexed leader regardless of treatment. Taking boys and girls together, the effectiveness of the treatments was shown in Table 6.

5. Initial level of performance of *Ss*: In the color matching data, levels of initial performance interact significantly with the sex of the *Ss* ($p < .01$). In the learning by imitation aspect of the study the effect of initial level of performance was significant at the .07 level. Initial level of performance then affects behavior in both types of imitation, but the direction of the effects are not the same. One might say that the brighter girls are more like the boys in that they improve more in the learning by imitation. Conversely, one might say that the brighter boys are more like the girls in that they color match more. Actually, bright boys, less bright girls, bright girls, and less bright boys color match in that order, whereas bright boys, less bright boys, bright girls, and less bright girls learn by imitation in that order (*see* Table 8).

TABLE 8

PERFORMANCE IN IMITATION AS A FUNCTION OF TYPE OF IMITATION, SEX OF SUBJECTS, AND INITIAL LEVEL OF PERFORMANCE

	BOYS		GIRLS	
Groups	*Same Color Choices*	*New Passes*	*Same Color Choices*	*New Passes*
	PERCENTAGE	AVERAGE NO.	PERCENTAGE	AVERAGE NO.
1 & 2	21	2.3	40	1.5
3 & 4	43	2.8	30	1.9

At this point we should return to the theoretical outlooks which were discussed in this paper and in "Learning by Imitation in Kindergarten Children" (5) and which were the origins of the variables investigated. We can now examine them in the light of the findings of these studies.

First, the reinforcement view of learning which Miller and Dollard hold led us to the expectation that the female should be the most effective model for imitation. This expectation was based on the assumption that five-year-old children (both boys and girls) have been reinforced for imitating more frequently by a female (the mother and now the teacher). Hence, a female should have more secondary reward properties and so elicit more imitative responses. This was not the case either for learning by imitation or for color matching. The author is aware of the possibility that the mother might reinforce the boy for imitating males rather than for imitating her. This, however, would not serve to explain why the female was not the most effective model for the girls for either type of imitation. Indeed, the only condition in which the female leader was highly effective was in the case of boys toward whom she had been attentive throughout the preliminary period.

Second, a general view of identification theory led to the expectation that identification should be sufficiently advanced by kindergarten age that an adult of the same sex as the child should be the most effective model for eliciting imitative responses. This was not generally the case. It *was* the case for boys in the color matching aspect of the experiment.

Third, the role of threat in leading to identification has been considered by a number of theorists. Their considerations led to the expectation that the mild threat involved in withdrawal of attention should lead to more identification and hence to a greater effectiveness of the model who used it, at least with respect to behaviors demanded or suggested by the model. Threat should be particularly effective when it comes from an adult of the same sex as the child. The contingency view of reinforcement might lead to the same expectation. However, in that case, the sex of the person who withdraws attention (threatens the relation) is of less obvious importance. From identification theory, the prediction is clearest for boys with an adult male leader. However, the findings of Hartup (1) led to the expectation that withdrawal of attention from girls by an adult female should be more effective in evoking learning by imitation.

For girls in the matched-dependent imitation study, withdrawal of attention was effective in the opposite way. There was less imitation found in girls who had a leader who withdrew attention. This was true regardless of the sex of the leader. The only counter tendency was for the male leader who withdrew attention to be effective in evoking learning by imitation.

For boys, both the color matching and learning by imitation experiments showed a tendency for withdrawal of attention from the same sexed leader to lead to more imitation. However, neither difference reached the usual level of statistical significance except for boys who were in the higher (brighter) groups according to initial level of performance on the mazes. With the opposite sexed leader, attention was more effective for learning by imitation. In color matching (which was relatively infrequent) attention or its withdrawal made no difference.

Fourth, consideration of the Miller and Dollard view that all forms of imitation are learned in the same way might lead one to expect both kinds of imitation to relate to the variables in the same way. As we have seen, sometimes they do and sometimes they do not.

Fifth, the author had hypothesized that learning by imitation (task oriented or "instrumental" imitation) might be more sensitive to the withdrawal of attention manipulation (threat), whereas matched dependent imitation ("role" oriented imitation) might be more responsive to attention (reward). The only real support for these notions lies in the data of the girls where attention seemed to be the crucial variable in determining the amount of color matching. There is a kind of counter evidence with boys. For them there was some tendency for threat to have an effect in enhancing imitation, and the effect was the same for both kinds of imitation.

In the light of these findings it appears that the variables we have selected for study do have effects. However, the variables interact, and their effects are not those to be expected from any one of the theoretical approaches considered here. In short, the relations of the variables to the amount of imitation are highly complex, and no existing theory is sufficiently differentiated to account for them. One might summarize loosely by saying that in this situation: For girls, color matching depends on attention from the model. Learning by imitation depends on having a male (opposite sexed) model. For boys, color matching depends on having a male (same sexed) model. Learning by imitation is more complex; it depends on having a same sexed leader who withdraws attention or an opposite sexed leader who is attentive. In all cases it is necessary also to take into account the initial level of performance or ability of the *Ss*.

It seems probable that each of these approaches is right in some circumstances and that almost all of them need to be taken into account in formulating a theory adequate to all circumstances. Replication and extension of this type of study appears to be a promising avenue for obtaining data on which a new formulation of the problems of identification and imitation can be based.

Summary

This study was carried out in order to test the hypothesis that matched-dependent imitation would respond to a particular set of experimental manipulations in the same way that learning by imitation or copying did. The experimental manipulations involved the sex of the adult leader to be imitated, the sex of the leader in relation to that of the child, and the way in which the leader treated the child in the 10-minute period preceding the imitation session (attentive throughout vs. withdrawing attention for half of the period). The matched-dependent behavior under study involved color matching. A child could match the color of pencil used by an adult to do the Porteus mazes. There was a pretest session in which the child's color preferences were established in addition to his level of performance on the mazes. There was then an experimental (or copying) session in which the child observed an adult before trying the task. A datum which could be observed by the child (other than the way in which the adult performed on the maze) was the color of pencil which the adult used in doing the maze. In the experimental sessions the actual work on the mazes was preceded by a 10-minute period in which the adult behaved in one of two ways: attentively throughout, or attentively for half the time followed by withdrawal of attention for the remaining half of the time.

The data showed a number of trends in the direction of supporting the above hypothesis, and some contradictory to it. Supplementary evidence showed that attention vs. withdrawal of attention makes a difference in the amount of influence a leader has in changing the child's behavior. There

were a number of interesting differences between the two types of imitation that need to be explored further.

References

1. Hartup, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. *Child Develpm.*, 1958, 29, 191-201.
2. Miller, N. E., & Dollard, J. *Social learning and imitation.* Yale Univer. Press, 1941.
3. Porteus, S. D. *The Porteus Maze Test and intelligence.* Pacific Books, 1950.
4. Rosenblith, J. F. Imitation in kindergarten children. Unpublished doctoral dissertation, Radcliffe College, 1958.
5. Rosenblith, J. F. Learning by imitation in kindergarten children. *Child Develpm.*, 1959, 30, 69-80.
6. Schein, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, 49, 389-395.
7. Schwartz, N. An experimental study of imitation: The effects of reward and age. Unpublished senior honors thesis, Radcliffe College, 1953.

Notice to Contributors

Child Development is devoted to basic research and theory concerning the growth and development of the child.

Manuscripts should be submitted in duplicate, each copy bearing only the title as a means of identification. The title, author's name, and institutional affiliation (together with any footnotes which reveal the identity of the author) should be submitted on separate pages.

Contributors are asked to refer to the *Publication Manual* of the *American Psychological Association* (1957 Revision) in the preparation of manuscripts.

The following rules require special emphasis:

1. All copy, including indented matter, should be typed double-spaced on white standard paper. Lines should not exceed six inches in length.

2. Footnotes should be few in number. They should be indicated in the text by superscript arabic numerals, numbered consecutively throughout the manuscript. The footnotes themselves should be typed double-spaced on a separate page placed at the end of the manuscript.

3. Tables and figures are costly to print and should be used only when essential. The author will be asked to contribute to the cost of composition of tabular and graphic material which is excessive in amount or unduly complicated in form. Each table should be typed on a separate sheet. Insert a guide line (e.g., Table 1 about here) at the appropriate place in the manuscript. Figures should be prepared with black India ink, suitable for photography. Legends typewritten on an ordinary typewriter are unacceptable.

4. References in the text should be indicated by numbers in parentheses that refer to the list of references at the end of the article. Do not use footnotes for this purpose. In the list of references, the arrangement should be alphabetical by author's surname, or for multiple authors by the name of the senior; for any other author, by order of date. The following standard form should be observed:

Jones, H. E. Development in adolescence. D. Appleton-Century, 1943.

Shepard, W. O., & Schaeffer, M. The effect of concept knowledge on discrimination learning. Child Develpm., 1956, 27, 173-178.

The World List of Abbreviations is used for journals.

The manuscript submitted should represent the final form in which the author wishes the paper to appear. All changes in proof (other than printer's errors) will be charged to the author.

All manuscripts should be addressed to William E. Martin, Editor, Child Development Publications, Purdue University, Lafayette, Indiana.

CONTENTS

(continued from front cover)

Child Behavior Problems and Parental Attitudes
Donald R. Peterson, Wesley C. Becker, Donald J. Shoemaker, Zella Luria, and Leo A. Hellmer 151

Discrimination Performance as a Baseline for Individual Analysis of Young Children
Sidney W. Bijou 163

Cultural Stereotypes and the Validity of Interviews for Research in Child Development
Joan McCord and William McCord 171

Effects of Maternal Employment on the Child
Lois Wladis Hoffman 187

Test Anxiety and Classroom Observations
Kenneth S. Davidson and Seymour B. Sarason 199

Imitative Color Choices in Kindergarten Children
Judy F. Rosenblith 211

CURRENT MONOGRAPHS

of the Society for Research in Child Development, Inc.

Helen L. Koch
THE RELATION OF CERTAIN FORMAL ATTRIBUTES OF SIBLINGS TO ATTITUDES HELD TOWARD EACH OTHER AND TOWARD THEIR PARENTS 1960 $3.50

Nancy E. Wood
LANGUAGE DEVELOPMENT AND LANGUAGE DISORDERS: A COMPENDIUM OF LECTURES 1960 $2.75

Irwin M. Marcus, *et al.*
AN INTERDISCIPLINARY APPROACH TO ACCIDENT PATTERNS IN CHILDREN 1960 $2.50

Bernard Farber
FAMILY ORGANIZATION AND CRISIS: MAINTENANCE OF INTEGRATION IN FAMILIES WITH A SEVERELY MENTALLY RETARDED CHILD 1960 $2.75

(A mimeographed list of all monographs available in the series since 1935 will be sent upon request.)

order from

CHILD DEVELOPMENT PUBLICATIONS

Purdue University, Lafayette, Indiana